Deep learning for industrial image: challenges, methods for enriching the sample space and restricting the hypothesis space, and possible issue

Article *in* International Journal of Computer Integrated Manufacturing · March 2021

DOI: 10.1080/0951192X.2021.1901319

Tianyuan Liu
The Hong Kong Polytechnic University

Bao Jinsong
Donghua University

Junliang Wang
Donghua University

INTERNATIONAL JOURNAL OF COMPUTER INTEGRATED MANUFACTURING
https://doi.org/10.1080/0951192X.2021.1901319

Taylor & Francis
Taylor & Francis Group

ARTICLE

Deep learning for industrial image: challenges, methods for enriching the sample space and restricting the hypothesis space, and possible issue

Tianyuan Liu, Jinsong Bao, Junliang Wang and Jiacheng Wang

College of Mechanical Engineering, Dong Hua University, Shanghai, China

ABSTRACT
Deep learning (DL) is an important enabling technology for intelligent manufacturing. The DL-based industrial image pattern recognition (DLBIIPR) plays a vital role in the improvement of product quality and production efficiency. Although DL technology has been widely used in the field of natural image, industrial image often has some mixed characteristics, such as small sample, imbalance, small target, strong interference, fine-grained, temporality and semantical, which reduce the feasibility and generalization of DLBIIPR. To solve this problem, this paper provides an overview of approaches commonly used in industry by enriching the sample space and limiting the hypothesis space. In order to improve the confidence of front-line workers in using DL models, the explainable deep learning (XDL) methods are reviewed, and a case study is used to verify the effectiveness of XDL.

ARTICLE HISTORY
Received 1 July 2020
Accepted 3 March 2021

KEYWORDS
Intelligent manufacturing; deep learning; industrial Image; pattern recognition; over-fitting; computational learning theory; sample space; hypothesis space; explainable deep learning

1. Introduction

1.1. Definition of industrial image and industrial image pattern recognition

Intelligent manufacturing (IM) is the unified goal of the next generation of manufacturing strategies in various countries. In order to realize IM, a key step is to collect information of the physical world through sensors. The industrial manufacturing process involves multiple stages such as design, manufacturing, testing, packaging, and maintenance. There are many sensor sources which exist in different stages. Due to the fact that 80% of the information humans obtain from nature comes from vision, vision sensors are used extensively in industry. The vision sensors convert the optical signal in the industrial manufacturing process into an analog current signal through the photosensitive element, and the current signal is amplified and digital-to-analog converted to obtain a digital image. Although there is no clear definition of industrial image, the above analysis reveals that the basic connotation of industrial image is a visual signal that can reflect the semantics of industrial manufacturing under the constraints of specific manufacturing rules.

After acquiring industrial image, the industrial image pattern recognition (IIPR) technology can analyze the brightness, color and pixel distribution information contained in the industrial image to reflect the characteristics of the industrial manufacturing process such as category, size, shape, etc. The IIPR task can be divided into four levels according to the actual requirements: know-what, know-where, know-how, know-when-and-where. The first level of IIPR task needs to answer what is the target category in the industrial image. The second level of IIPR task to address where the target is located in the industrial image. The third level of the IIPR task needs to address what is the size of the target in the industrial image. The fourth level of the IIPR task needs to address where the target appears at what moment in the industrial image. It is worth noting that the four levels of tasks mentioned above are not completely independent. For example, the tasks at the second and third levels generally have to include the task at the first level as well. Therefore, the task at the first level is the most fundamental one.

1.2. Current development of IIPR

The vision sensors and the pattern recognition modules constitute the vision system. The visual systems are essential for people to gain insight into the industrial production process. Vision systems are widely

CONTACT Jinsong Bao bao@dhu.edu.cn College of Mechanical Engineering, Dong Hua University,Shanghai 201620, China.

used in metal manufacturing (steel, welding, casting, etc.) (Zhao, Peng, and Yan 2018; T. Lei et al. 2020; W. Z. Du et al. 2020), railway manufacturing (rail, fastener, crimping part, etc.) (Y. P. Wu et al. 2020; Wei et al. 2019; Q. Han et al. 2020), textile manufacturing (yarn, fabric, garment, etc.) (Xiang et al. 2019; H. Zhang et al. 2019; Donati et al. 2019), electronic component manufacturing (wafer, circuit board, display screen, etc.) (Wang and Chen 2019; Hassanin, Fathi, and Ghada 2019; Jian, Gao, and Ao 2017), electrical manufacturing (photovoltaic, insulator, cable, etc.) (Li et al. 2020; A. Ibrahim et al. 2020; Michalski et al. 2019), etc. Reliable and efficient IIPR are the key to the vision system.

The development of the industrial vision system mainly gone through three stages as shown in Figure 1. The first stage acquires visual signals through the eye and then the signals are processed by the brain. Although there is a higher degree of flexibility at this stage due to human involvement, efficiency and quality consistency are comparatively low. The second stage is the traditional machine learning-based IIPR. This stage mainly follows the process from image pre-processing to feature extraction and then to feature selection and feature recognition. The image pre-processing stage mainly involves operations such as filtering (Feng et al. 2020), denoising (Malarvel et al. 2017), enhancement (Lin, Wu, and Hong 2012), and edge detection (Yu et al. 2017). The primary geometric features in industrial images (Valavanis and Kosmopoulos 2010) or feature descriptors represented by SIFT (Yang et al. 2020), HOG (Ming et al. 2019), and LBP (Zhou et al. 2020) are mainly extracted in the feature extraction stage. The main methods involved in the feature selection stage are principal component analysis (Z. Y. He et al. 2019), independent principal component analysis (Ahmad et al. 2019), and sequential forward floating selection (Wang et al. 2017). The main methods involved in the feature recognition stage are perceptron (Zeng, Dai, and Mu 2007), support vector machine (Xiao et al. 2020), decision tree (Z. F. Zhang et al. 2020), and clustering (Wang et al. 2015).

This stage is a process of human-driven feature design, extraction and selection, so the model is more explainable and reliable in some fixed scenarios. However, too many intermediate steps would lead to the following deficiencies. First, the outcome of each intermediate task will affect the final result, and the optimal solutions of the subtasks does not imply the optimal solution of the overall problem. Second, each step is a separate subtask, which will severely affect the overall efficiency of IIPR and makes it difficult to adapt to the complex and changing production environment. Third, the limited priori knowledge of human will result in the data not being fully valued. The third stage is the deep learning (DL) -based IIPR (DLBIIPR). The main idea at this stage is to select a certain hypothesis function from the hypothesis space as the final model according to the learning algorithm after acquiring the industrial image dataset. The DL-

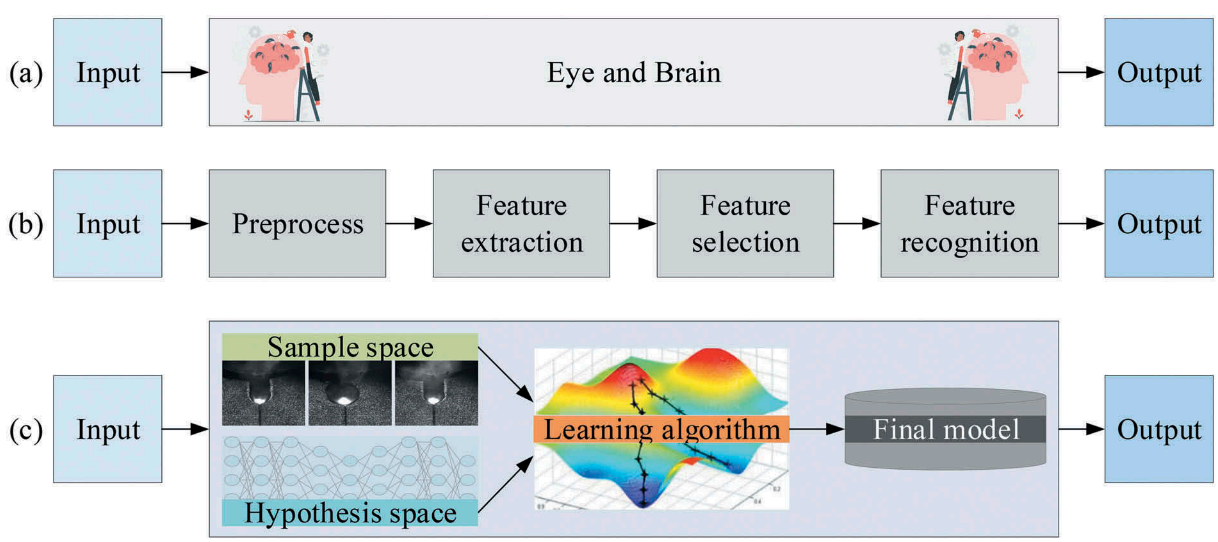

Figure 1. The development of the visual system. a) Human-based; b) Traditional machine learning-based; c) DL-based.

based methods can automatically learn deep abstract information to better represent image patterns and this end-to-end modeling approach has greatly improved the efficiency of IIPR. Under the general trend of the continuous integration of intelligent technology into the manufacturing industry, the DLBIIPR is current research hotspot.

1.3. Challenges for DLBIIPR

Although DL has been widely researched and applied in pattern recognition of natural image, industrial image has some characteristics different from natural image due to the particularity of the manufacturing process. These characteristics lead to the limited application of DLBIIPR. Figure 2 shows six typical public datasets in industry. It can be found that industrial image mainly has these following characteristics. **A) Small sample**. First, since IM strategies are just beginning to be used, the industry's attention and accumulation of data are just beginning to increase. Second, due to the relative maturity of the production process, samples of defects do not appear frequently. Third, due to the large number of industrial subdivisions, the labeling of data is a professional and time-consuming process, which results in less data labeled. **B) Imbalance**. As mentioned above, defective samples do not occur frequently in industrial production, which leads to a significant imbalance in the amount of data between normal samples and defective samples. Furthermore, the possibility of finding different categories of defects is also various, which leads to the imbalance between different categories of defects. In addition, the feature patterns of different categories are different, which leads to the imbalance of the learning ability of the same deep learning model (DLM) for different categories (Wang et al. 2020a). **C) Strong interference**. The industrial environment is complex, and the production process generally involves complex physicochemical reactions, which are accompanied by changes in light, vibration, motion blurring, etc. These disturbances seriously

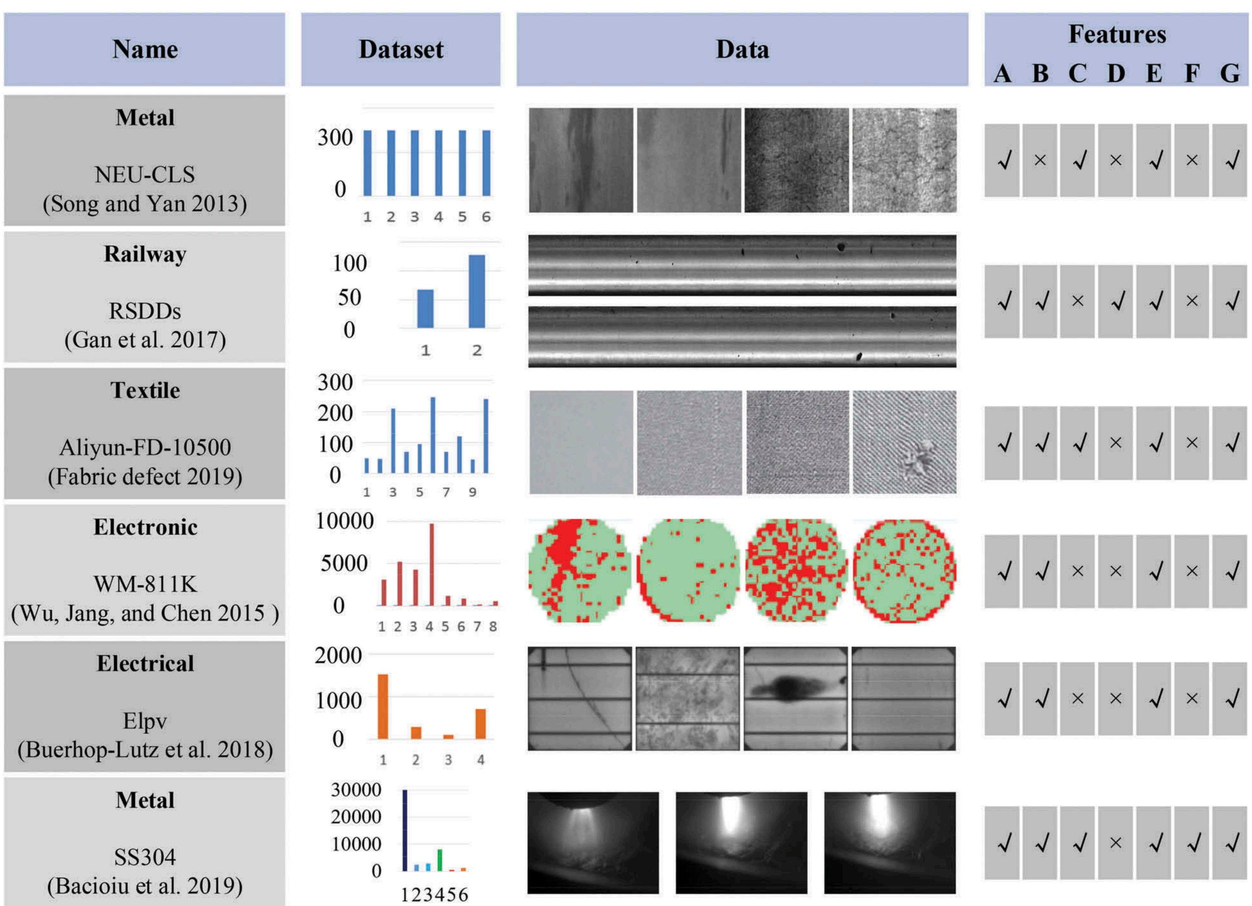

Figure 2. Typical industrial image datasets and characteristics.

hinder the mining of effective patterns by DLMs. **D) Small object**. Industrial products are generally made up of many small components, and defects in industrial products are generally relatively smaller than the product itself. The underlying backbone neural networks (VGG and Resnet) of the DLMs have multiple down-sampling processes. The size of the small targets in the feature map is basically only single-digit pixel size, which severely limits the descriptive effect of the DLMs on small targets. **E) Fine grained**. Since image analysis in industrial production is generally applied in a subdivision field, this will result in little macroscopic difference between different images. However, there are great differences within the same category due to the existence of occasional circumstances. It used to take skilled workers to pick out subtle features. This is a big challenge for DLMs, which must find the most discriminating abstract features between fine-grained images. **F) Temporality**. The manufacturing stage is generally a continuous process, and complex industrial manufacturing features often need to be mined with timing information. This puts forward requirements for the complexity of the DLMs. **G) Semantical**. The industrial manufacturing process is carried out under clear production rules, so the industrial image contains clear industrial manufacturing semantics. This kind of industrial manufacturing semantics is vital to the traceability of industrial manufacturing processes. However, the current DLMs can explain poorly and have insufficient ability to reflect industrial semantics. Even more unfortunate is the fact that in practice the above issues are often combined.

It can be seen from Figure 1-c that DLBIIPR mainly involves four aspects: sample space, hypothesis space, learning algorithm and final model. Since deep learning generally uses gradient descent as the learning algorithm, the research involved in DLBIIPR mainly focuses on the sample space, hypothesis space and the final model. A comparative analysis of the characteristics of industrial image reveals that small sample and imbalance belong to the characteristics of industrial image dataset. The characteristics of strong interference, small object, fine grained, temporality and semantical are specific to industrial image itself. From the perspective of deep learning, the challenges caused by small sample and imbalance are mainly for the sample space. The challenges brought by strong interference, small object, fine grained, and temporality are mainly for the hypothesis space. The challenges brought by industrial image with specific industrial manufacturing semantics are mainly for the final model.

Based on knowledge of computational learning theory, features from the industrial image dataset raise the issue that the empirical error of the DLMs can converge to 0, but the difference between the generalization error and the empirical error is relatively large. The characteristics of the industrial image itself will make it difficult for the empirical error of the DLMs to approach 0, and the generalization error is still far from the empirical error. Furthermore, it is known that the characteristic of small sample can lead to insufficient information in the sample space, such that DLMs lack inductive bias. The unbalanced feature of the dataset leads to inductive bias that are far from the actual problem. The characteristics of the industrial image itself will lead to the complexity of the image patterns and therefore require a larger space of hypothesis to find an optimal solution. In addition, the explicit semantic nature contained in industrial image contradicts the unexplainability of DLMs.

The problems from the three aspects of sample space, hypothesis space and final model will eventually lead to the poor feasibility and generalization of DLBIIPR. Among them, the existence of problems in the sample space and hypothesis space can cause the learning process to become unreliable, which subsequently leads to the poor feasibility and generalization ability of DLBIIPR. The poor feasibility and generalization ability caused by the unexplainability of the model is more from the application perspective, especially for security-sensitive application scenarios.

1.4. Overall idea

At present, there is currently no review article on methods to improve the feasibility and generalization ability of DLBIIPR. Therefore, this paper will provide an overview of the methods in terms of enriching the sample space, limiting the hypothetical space and enhancing explainability in order to provide a reference for industry personnel to better apply DL methods for IIPR. The general idea is shown in Figure 3. It should be noted that the three aspects are carried

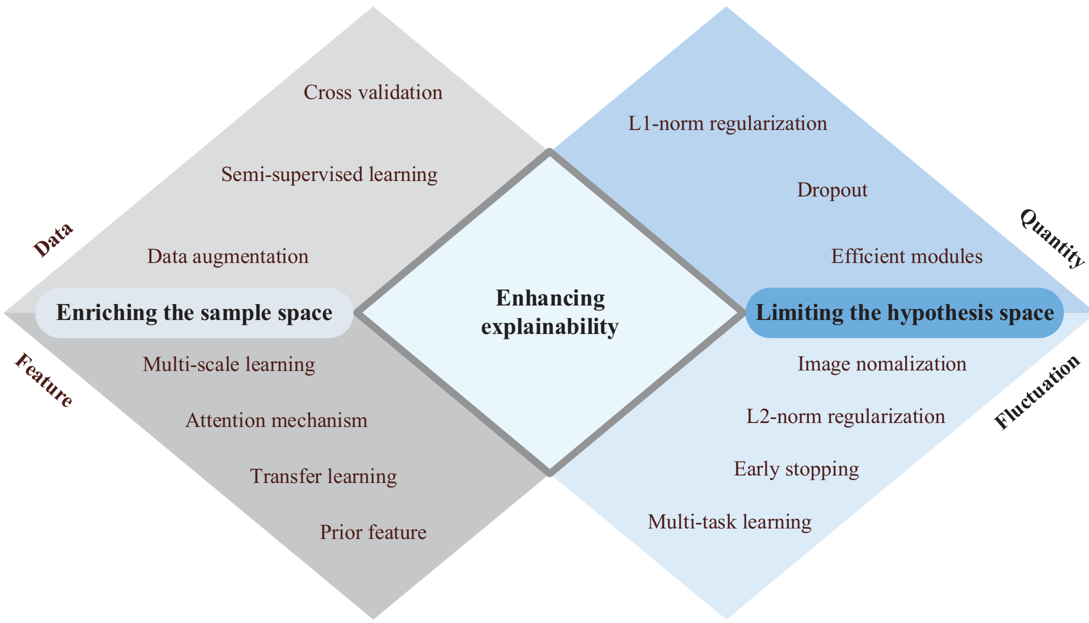

Figure 3. The overall idea of this article.

out to facilitate the organization of the review, but in practical applications, just as the problems in industrial image often appear in combination, various methods are also used in combination. In addition, since explainable methods are just emerging in the field of deep learning, there are not yet many applications and studies of explainable methods in industry. Therefore, the review of model explainability is presented as future challenges and possible issue in this paper. Firstly, the current industrial requirements are analyzed, and then the explainability methods in the field of deep learning are outlined. Finally, a case study is used to verify the effectiveness of the explainability methods in the field of IIPR. The explainability methods can help developers to improve the deep learning model in a targeted manner. This improvement is essentially to standardize the training direction of the network and the function of the model by increasing the interaction between developers and the deep learning models. Therefore, the introduction of explainability methods will not only reduce the size of the hypothesis space, but also increase the possibility of DLMs being applied in safety-sensitive industrial scenarios.

1.5. Contributions

This paper makes the following contributions: **A)** The characteristics of industrial images are analyzed. This paper analysis the underlying reasons that these characteristics restrict the application of the DLBIIPR from the perspective of computational learning theory. **B)** From the perspective of 'enriching the sample space', the common methods in the DLBIIPR are reviewed. The advantages and disadvantages of each method are analyzed. Furthermore, the future research direction is proposed. **C)** From the perspective of 'limiting the hypothesis space', the common methods in the DLBIIPR are reviewed, the advantages and disadvantages of each method are analyzed, and the future research direction is proposed. **D)** The explainable deep learning (XDL) methods are reviewed, and a case study is used to verify the effectiveness of XDL.

2. Related foundations

2.1. Deep learning

2.1.1. Concept and a brief history of deep learning

Geoffrey Hinton and his student Ruslan Salakhutdinov formally introduced the concept of DL in 2006 (Hinton, Osindero, and The 2006). In 2012, the team led by Jeffrey Sinton won the championship with AlexNet in the famous Imagenet image recognition competition, which has opened a new era of DL (Alex, Sutskever, and Hinton 2012). In 2016, the boom about deep learning was pushed to new heights with AlphaGo's 4:1 win over top international Go player Li Shishi.

DL methods are representation-learning methods with multiple levels of representation, obtained by composing simple but non-linear modules that each transform the representation at one level (starting with the raw input) into a representation at a higher, slightly more abstract level (Lecun, Bengio, and Hinton 2015). The definition of DL is analyzed and summarized in the literature (W. J. Zhang et al. 2018). DL is a process not only to learn the relation among two or more variables but also the knowledge that governs the relation as well as the knowledge that makes sense of the relation. Typical DLMs conclude convolutional neural network (CNN), recurrent neural network (RNN), auto encode and deep belief network (J. J. Wang et al. 2018).

Since images are typical grid-type data and have many close spatial context connections, the CNN framework is the most widely used in image processing tasks. Since Lecun proposed lenet-5 in 1998 to indicate the beginning of CNN (Lecun et al. 1998), the championship performance of Alexnet in 2012 has brought CNN back into people's vision and causes a stir. After that, the classical CNN infrastructure emerged. The development of CNN-based DLMs in image processing related tasks is shown in Figure 4. In terms of basic classification tasks, typical structures include VGG16 (Simonyan and Zisserman 2015), Googlenet (Szegedy et al. 2015), Resnet (K. M. He et al. 2016), Mobilenet (Howard et al. 2017), SEnet (J. Hu et al. 2019), etc. VGG16 proves that increasing the depth of the network can affect the final performance to a certain extent. Googlenet focuses on extracting multi-scale features in the width direction of the network by integrating multi-dimensional convolutional kernels. Resnet makes it possible for networks to increase in depth without degradation through the learning of constant mapping. Mobilenet's deeply separable convolution method significantly reduces the number of parameters and the number of operations in the convolution process. SEnet adds a module to model the importance of feature channels. Through this development process, it can be found that the basic CNN has conducted a comprehensive research on the depth, width, lightweight and efficient modules of the network. As shown in Figure 4, with the development of CNN infrastructure, many achievements have also been made in tasks such as target detection (Girshick et al. 2014; Girshick 2015; Ren et al. 2016; Redmon et al. 2016; Liu et al. 2016), image segmentation (Shelhamer, Long, and Darrell 2017; Ronneberger, Fischer, and Brox 2015; Kendall,

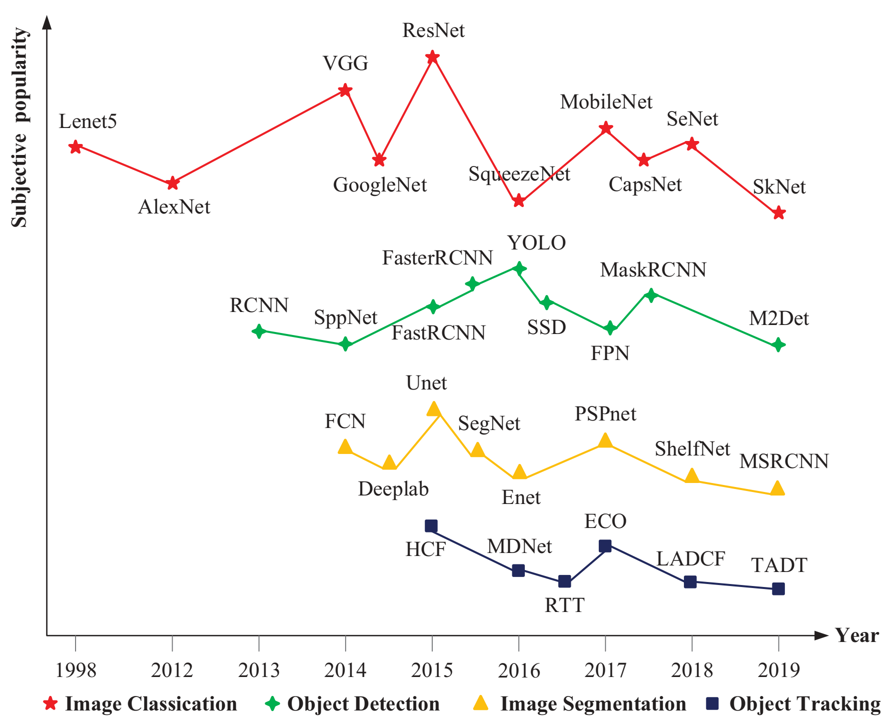

Figure 4. The development of CNN in different tasks.

Cipolla, and Cipolla 2017; Zhao et al. 2017; Zhuang and Yang 2018.), and target tracking (Ma et al. 2015; Nam and Han 2016; Danelljan et al. 2017; Xu et al. 2019; Li et al. 2019). For example, CNN is used to extract features of candidate regions for target detection and CNN is used to extract image features for pixel-by-pixel classification for image segmentation.

2.1.2. Deep learning, machine learning and artificial intelligence

The concepts that are easily confused with DL include artificial intelligence (AI) and machine learning (ML). AI is a technical science that researches and develops theories, methods, technologies and application systems for simulating and expanding human intelligence (Shehab 2020). The object of its research and development is 'theory, technology and application system', and the purpose of the research is to 'simulate and expand human intelligence'. The most basic approach of ML is to use algorithms to analyze data and then make decisions about events in the real world. Machine learning methods have a wide range of application scenarios in many industrial manufacturing departments such as design (Cui et al. 2020), manufacturing (Tian et al. 2021), inspection (Guan et al. 2021) and management (Lu et al. 2019, 2017). DL is a kind of ML technology, which is used to build and simulate the neural network of human brain for analysis and learning, and simulate the mechanism of human brain to interpret data. Its basic feature is to use a deeper network to mimic the mode of transferring and processing information between neurons in the brain.

Generally, AI is a goal in terms of logical relationships, while ML is a technical means to achieve this goal, and DL is a kind of method in ML that has greater ability to express data (Alexopoulos, Nikolakis, and Chryssolouris 2020). Thus from the perspective of inclusion relationship, the relationship among AI, ML and DL is shown in Figure 5.

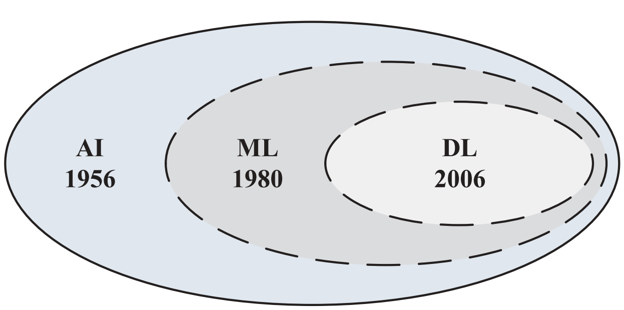

Figure 5. The relationship between AI, ML and DL.

2.2. Computational learning theory

Computational learning theory examines the theory of learning through computation. It aims to analyze the difficult nature of the learning tasks and provide theoretical assurance on the feasibility of the learning algorithm and to guide the design of the algorithm. The feasibility of DL algorithms can be discussed in the following three situations: the hypothesis space has only one hypothesis function, the hypothesis space has a limited number of hypothesis functions, and the hypothesis space has an infinite number of hypothesis functions (Zhou 2016).

- Only one hypothesis function

The proportion of prediction error on the unknown sample is recorded as $E_{predict}$ and the proportion of training error on the known sample is recorded as E_{train}. The unique hypothetical function is recorded as h^*, e is the set tolerance error range, and N is the sample size. When the following relationship is satisfied:

$$P\left(\left|E_{predict}(h^*) - E_{train}(h^*)\right| \geq e\right) \leq 2e^{-2Ne^2} \quad (1)$$

It is said that $E_{predict} = E_{train}$ is Probably Approximate Corrent (PAC). When N is large enough, the probability that the difference between training error" and 'prediction error' exceeds the error range will be very small, so learning is feasible. When the learning algorithm has no choice, it is almost impossible to obtain a model with a high training accuracy. Although the prediction accuracy at this time can reach a level similar to the training accuracy, it is likely that the performance is same poor rather than same good.

- A limited number of hypothesis functions

Let the hypothesis function that makes $\left|E_{predict}(h^*) - E_{train}(h^*)\right| \geq e$ valid be a bad hypothesis. When there are M hypothesis functions in the hypothesis space, any hypothesis function is a bad hypothesis. Because any of the hypothesis functions can be selected by the learning algorithm as the final model.

The probability of a bad hypothesis in the hypothesis space is as follows:

$$P(badhypothesis) \leq P(h_1\ isbadhypothesis) + P(h_2\ isbadhypothesis) + \cdots\cdots + P(h_M\ isbadhypothesis)$$

$$\leq 2e^{-2Ne^2} + 2e^{-2Ne^2} + \cdots\cdots + 2e^{-2Ne^2} = 2Me^{-2Ne^2} \tag{2}$$

As can be seen from the formula (2), when there are M hypothesis functions in the hypothesis space, the probability of a bad hypothesis appearing becomes M times that when there is only one hypothesis function. However, as long as the sample size N is large enough, it is still guaranteed that the probability of bad hypotheses appearing is very small. That is, $E_{predict} = E_{train}$ is still PAC. Moreover, due to the increase of hypothesis functions, the learning algorithm has room for optimization. Compared with the hypothesis space where there is only one hypothesis function, it is more likely to learn a model with high training accuracy.

- An infinite number of hypothesis functions

Take the dichotomy task as an example. In this case, although there are an infinite number of hypothetical functions, there are only positive or negative results for a given sample of classification. Taking this idea as the core, Vapnik and Chervonenkis proposed and proved the VC-bound theory in 1971. According to this theory, the following formula can be derived.

$$P\left(\left|E_{predict}(h^*) - E_{train}(h^*)\right| \geq e\right) \leq 2\frac{2\sum_{i=0}^{VCD}\binom{2N}{i}}{e^{\frac{2}{16}Ne^2}} \tag{3}$$

In the formula (3), VCD represents the VC-dimension of the hypothetical space. According to this formula, we can know that when there are an infinite number of hypothesis functions in the hypothesis space, the probability of occurrence of a bad hypothesis can still be limited by the sample size N. That is, $E_{predict} = E_{train}$ is still PAC.

It can be seen from the first case that when the hypothesis space is too small, complex image patterns would be often not discovered. And the larger hypothesis space has a stronger ability to establish complex relationships. However, it can be found that the larger the VCD, the larger the value to the right of the inequality. A larger sample size of N is needed if the probability of a bad hypothesis emerging is also to be maintained. Therefore, the complexity of the hypothesis space in practical application needs to be balanced by the size of the sample.

3. Enriching the sample space

Samples are the cornerstone of DL. In a broad sense, enriching the sample space is to expand the information contained in it. Enriching the sample space is the fundamental method to improve the feasibility and generalization ability of DLBIIPR. The current studies have mainly enriched the sample space in terms of sample size and sample features, so that the DLMs can own a more clear and reasonable inductive preference.

3.1. Cross validation

In the process of DL, the dataset is generally divided into three parts: the training set, the validation set and the testing set. The validation set is mainly used to observe the training process to optimize the hyperparameters. There are three types of cross validation (Arlot and Celisse 2010) in deep learning: k-fold cross validation, leave-one-out cross validation, and stratified k-fold cross validation. k-fold cross-validation method is the most commonly used in industry. As shown in Figure 6, k-fold cross validation generally divides the original dataset into k sub-samples. A separate sub-sample is reserved for model validation, and the other sub-samples are used for training. Each sub-sample will be used for validation, and a single estimate can be obtained ultimately by averaging the k results. The advantage of this method is that the limited sample space can be used as much as possible and data interference is eliminated to the greatest extent.

In the work by Xu et al. (2020), 3-fold cross validation is used to ensure that the same images do not appear in both the training and test sets. 4-fold cross validation is used in reference (Y. X. Lei et al. 2020). 5-fold cross validation is also used in reference (Ming et al. 2019; Kang 2020). What's more, the area under the receiver operating characteristics curve is used as the evaluate metric on the validation set (Kang 2020).

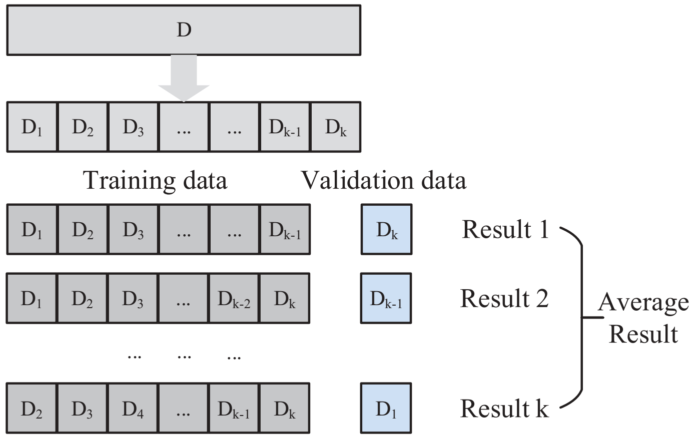

Figure 6. *K*-fold cross validation.

10-fold cross validation is used in reference (J. L. Wang et al. 2019) to ensure robust measures of validation performance. In (Scime and Beuth 2018), the authors partitioned the dataset according to 8:1:1 for 10-fold cross-validation.

However, studies on cross validation in DLBIIPR still have the following shortcomings. First, *k* is a hyperparameter, and the current study relies on experience for the choice of *k* and lacks optimal selection. Second, the *k*-fold cross validation does not take into account the original proportion of the data distribution when dividing the sub-dataset, so a stratified cross-validation method that incorporates the data distribution needs to be considered in future applications. Third, cross validation is only to maximize the use of limited sample space. But it does not introduce new information into the sample space. Therefore, the promotion effect of this method is very limited. Fourth, *k*-fold cross-validation means that the model has to be trained *k* times, which is *k* times longer than the traditional process.

3.2. Semi-supervised learning

Since the labeling of industrial images requires professional knowledge, it is difficult and costly to obtain large-scale labeling data. Although this problem severely limits the application of DLBIIPR, this problem also reveals that there are many unlabeled samples in practice. The semi-supervised learning method using labeled and unlabeled samples for model training provides a feasible way to solve this problem. The labeled data is used for model training, and the unlabeled samples can be used to improve model performance. As shown in Figure 7, there are five main categories of semi-supervised learning methods in

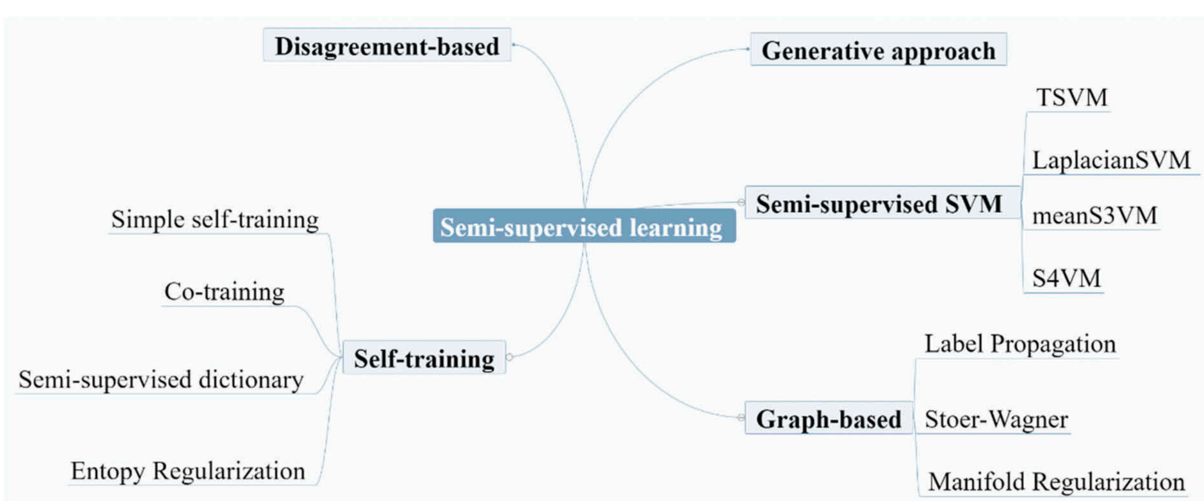

Figure 7. Classification of semi-supervised learning.

DL. However, there are not many applications of semi-supervised learning methods in the field of IIPR.

In (Gao et al. 2020a), the CNN is improved by pseudo-label (Lee 2013), which is an effective semi-supervised framework to generate fake labels for the unlabeled samples. This is a simple self-training method. The improvement of this study is in setting a parameter to balance the effect of unlabeled samples on the loss function. In (Lei et al. 2020b), Laplace regularization is used to match the local geometric information, thus fully mining the information contained in all available data. This is a method of manifold regularization. The limitation of this method is that it requires the assumption that labeled and unlabeled data are on the same low-dimensional manifold. In (Y. He et al. 2019aa), generative adversarial network (GAN) (Goodfellow et al. 2014) is used to generate label-free samples, and then assigned labels to the generated samples using discriminators and Resnet trained on the raw data. This is a method of co-training. Since GAN also requires labeled samples for training, the effectiveness of the method is strongly influenced by the GAN. In (D. He et al. 2019bb), a new semi-supervised learning method is proposed. In the first stage, convolutional auto-encode (CAE) is trained with unlabeled data. In the second stage, the encoder network of CAE is reserved as feature extractor and is fed into a softmax layer to form the discriminator. Then, GAN is introduced to generate fake images of steel surface defects to train the discriminator.

Compared with supervised learning, the unlabeled samples used in semi-supervised learning can improve the generalization ability on unknown data. But the semi-supervised learning algorithm has many requirements on the training data. Generally, the following points are assumed. The distribution of unlabeled data should be the same as that of labeled data. The labels of the labeled data should be correct. Unlabeled data is generally category-balanced. The true labels of unlabeled data are generally of a certain category in the labeled data.

3.3. Data augmentation

Data augmentation is the most fundamental and direct approach to the problem of insufficient information on sample space. As shown in Figure 8, the image data augmentation methods commonly used in DL are divided into two main categories: basic image processing methods and DL-based methods. The basic image processing methods include operations in geometric space and pixel space.

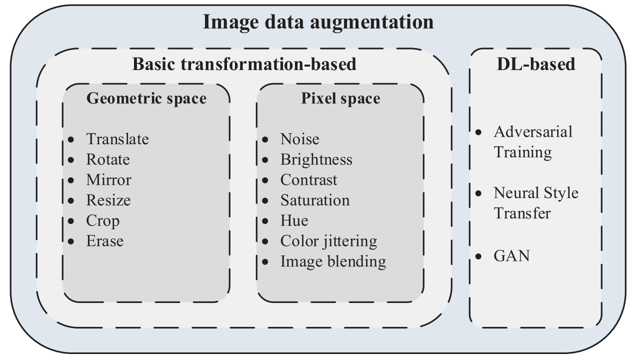

Figure 8. Methods for image data augmenting.

- In terms of basic transformation-based

There are five basic data augmentation methods are used in (T. He et al. 2019), which comes in the following forms: (a) rotating the training images at 90°, 180°, and 270°; (b) cropping the training images randomly; (c) flipping the training images horizontally, vertically, and diagonally; (d) changing the brightness, contrast, saturation, or hue of the training images; and (e) adding Gaussian and salt and pepper noises to the images. The images are performed horizontal flip, rotation, random cropping and histogram equalization (W. Z. Du et al. 2019). In (Cheon et al. 2019), they tripled the amount of training image by rotating each input image a random degree in the range [0, 180] and flipping the input image along the horizontal and vertical axes.

- In terms of DL-based

In (Lin, He, and Sun 2019), a method of defect enhancement generative adversarial network is proposed to solve the problem of insufficient sample size, which can generate micro-crack defects with obvious defect characteristics and high diversity. In (Niu et al. 2020), a new generation method called surface defect-generation adversarial network is proposed to address the problem of insufficient sample and unbalanced distribution, which generates defect images using a large number of defect-free images from industrial sites. In (Wang et al. 2019a), a novel DLM called adaptive balancing generative adversarial network is proposed for the defective pattern recognition of wafer maps with imbalanced data. In (Tang et al. 2020), the joint approach of combining GAN model and conventional image processing algorithm for image augmentation is used, i.e. GAN model is firstly used to generate the images and then the conventional image processing is used to enlarge the size of the sample image dataset.

The basic transformation method and GAN-based DL method have been most widely used. The basic transformation method is easy to operate and has no requirements on the initial data volume. However, since most of the information added by this method comes from the linear combination of existing information, this method can provide limited information for the sample space. From an application point of view, the selection of the basic transformation method needs to be closely related to the existing data and the task target. In the case of the data characteristics in (Kim et al. 2019), although 'rotation' is one of the common option for data augmentation, it only yielded better results for 'Binary' and 'Trunc' images and worse results for 'Gray' and 'Zero' images. In (Zhang, Wen, and Chen 2019), different noises were added for different types of penetration. The GAN-based method can theoretically learn the distribution of the original data to generate new data. However, due to the limitation of the original sample volume, the learned original distribution will be unreliable. Therefore, the quality of new samples generated according to the learned model cannot be guaranteed absolutely. In particular, Geirhos et al. (2018) proposed to use style-transfer to perform data augmentation, which allows the network to pay more attention to the shape of objects rather than data distribution, thereby improving the accuracy and robustness of the model.

3.4. Multi-scale learning

The multi-scale representation of the image is equivalent to introducing multi-scale information into the sample space. The trained DLMs will have scale invariance, which plays an important role in improving the ability of DLBIIPR. As shown in Figure 9, the multi-scale of geometric space and the multi-scale of Gaussian space are commonly used in industrial images. Multi-scale features from geometric space can be obtained by inputting images of different

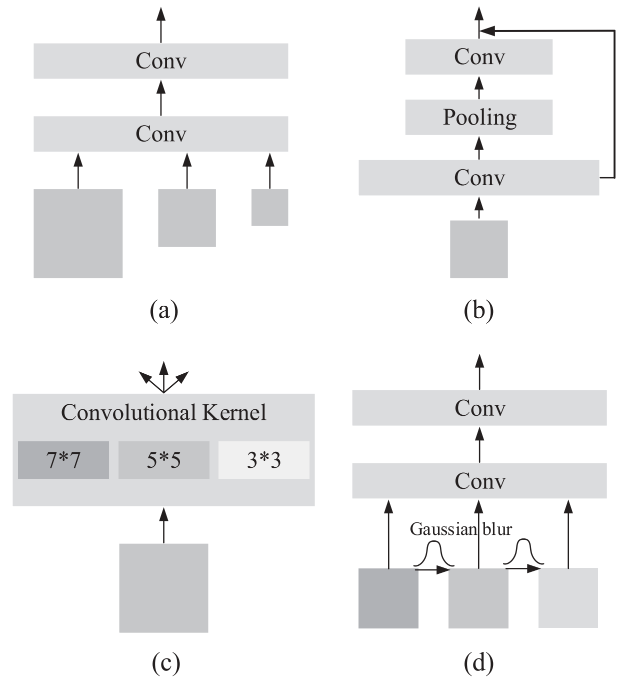

Figure 9. Multi-scale representations commonly used in industrial images. a) Different input resolutions; b) Pooling operation; c) Convolution kernels of different sizes; d) Gaussian kernels with different blurring levels.

sizes (Figure 9 (a)). Features with different resolutions can be obtained through pooling operations (or convolution operations) (Figure 9 (b)). The characteristics of different receptive fields can be obtained by convolution kernels of different sizes (Figure 9 (c)).

- In terms of geometric space

In (Scime and Beuth 2018), images of 25 * 25, 100 * 100 and 900 * 900 geometric scales are taken as different channels input for autonomous anomaly detection and classification in a laser powder bed fusion additive manufacturing process. This method can learn the key contextual information at multiple size scales. In (He, Xu, and Wang 2019a), convolution kernels of different sizes are used to extract features of different receptive field of the input image (different convolutional layers) to construct a multi-scale features to represent and classify the hot rolled defects. In (Y. He et al. 2019bb), convolution operation is used to obtain features of different resolutions, and multi-scale features are fused through region-of-interest pooling. The advantage of this pooling method is that there is no need to limit the size of the input.

- In terms of Gaussian space

The Gaussian pyramid is a simulation of the visual system of the human eye. In (Gao et al. 2020b; Mei, Wang, and Wen 2018), a three-level Gaussian pyramid is introduced to generate varying degrees of blur information of the defect, so that more information is available for model training.

In practical applications, extracting multi-scale features can certainly improve the accuracy of DLBIIPR. However, since it involves operations on multi-scale, further consideration is needed in terms of timeliness. In addition, the geometric scale and the Gaussian scale are basically considered in the application at present. Firstly, the study of the multi-scale mechanisms embedded in biological vision should also be given priority consideration. Zhao et al. (2020) were inspired by human vision and visual memory mechanisms. There are three categories of features are extracted, which are the visual perception information extracted by stacked convolutional auto-encoders, the visual short-term memory information characterized by a shallow CNN, and the visual long-term memory information characterized by non-local neural networks. Secondly, the multi-scale information of other domains should also be considered. In (J. Liang et al. 2017), four kinds of Gabor kernels are used to obtain the characteristics of multiple frequency bands. In (Y. P. Chen et al. 2019), adaptive methods are used to obtain high frequency and low-frequency features, respectively.

3.5. Attention mechanism

The attention mechanism emerged in imitation of the human way of thinking. The spatial pixels of the image are redundant expressions of objects, and the feature channels obtained through multiple convolution kernels are redundant expressions of features. For this problem, the attention mechanism can improve the accuracy and efficiency of DLBIIPR. As shown in Figure 10, the flexible attention mechanism based on location and channel is mainly used in the field of industrial images. The attention mechanism can be considered as adding priori information such as ‘the importance of different spatial positions in the image is not the same’ and ‘the importance of different channels in the feature tensor is different’ into the sample space. As a result, the DLMs will be trained in the direction of prior knowledge.

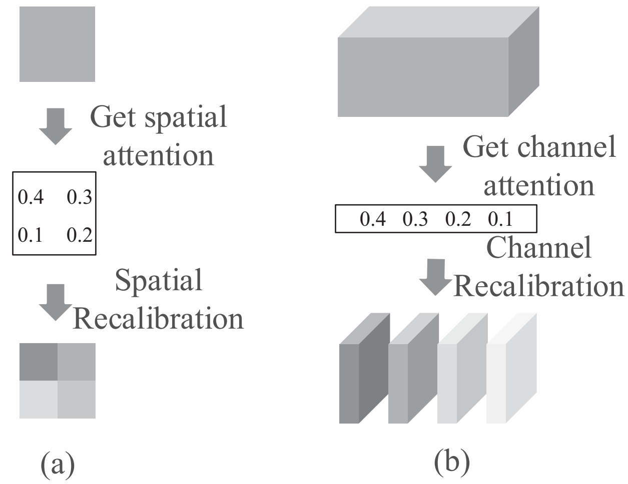

Figure 10. The attentional mechanism. a) Spatial attention; b) Channel attention.

The authors (X. Fu et al. 2020) propose a structure of attention gate. This structure uses convolution operation to obtain feature information of defective regions, and uses the Sigmoid activation function to generate attention values to suppress irrelevant regions. In (C. F. Hu et al. 2019), the authors inserted a spatial attention module after each block and visualized the attention map to demonstrate the inhibitory effect of the attention mechanism on the interference regions. In (Y. M. Wang et al. 2019), the soft attention template from the attention module is used as the weight of the feature map of the backbone network with convolution modules to improve the accuracy of defect detection. In (Xie et al. 2020; D. F. Zhang et al. 2019),dual-attention modules were introduced, i.e. attention was firstly applied to the feature channel and then to the location of the feature map. Thisdual-attention module has been proved by the authors (Woo et al. 2018) that the effect of applying channel attention firstly and then applying position attention is better.

The emergence of attention mechanism is inspired by cognitive science. Due to the bottleneck of information processing, human will selectively focus on a part of the overall information, and ignore other visible information. However, the current attention mechanism in DLBIIPR is mainly focus on the importance of spatial location and the importance of channels. From the perspective of neural connection, these two attention mechanisms are modeling the connection among neurons and the connection among feature maps. In addition, the deep neural network also has the connection among the feature layers. Therefore, this paper considers that the attention mechanism for the feature layers should be considered in the future.

3.6. Transfer learning

The connotation of transfer learning (Pan and Yang 2010) is shown in Figure 11. That is, in the case of a given source domain, source task, target domain and target task, it is a method to improve the target task by using some knowledge acquired from the source domain in solving the source task. Transfer learning relaxes the assumption that the training data must be independent and identically distributed (IID) with the testing data. The knowledge acquired from the source domain through transfer learning plays an expanding role in the target domain. There are four types of transfer learning methods in DL: instance-based, feature-based, model-based and relation-based. The most widely used in the industrial field is the feature-based transfer learning method. The feature-based transfer learning can be further divided into homogeneous transfer learning and heterogeneous transfer learning according to the similarity of data features between the source and target domains.

- In terms of heterogeneous transfer learning

The transfer learning method applied in industrial images is generally based on the VGG model. In the public data domain, Imagenet is generally selected as the pre-training data set. In (Gao et al. 2020b; X. Fu et al. 2020; Dung et al. 2019), the VGG16 network is pre-trained by Imagenet, and all the convolutional layers and max-pooling layers are retained. The top layer of VGG16 are fine-tuned. In (G. Z. Fu et al. 2019), the pre-trained SqueezeNet model is adopted as the backbone architecture, which contains nine fire modules. And the SqueezeNet model is pre-trained on the Imagenet dataset. A suitable model also needs to be matched with a suitable training process to get good results. Therefore, the ResNet model was firstly pre-trained using the Imagenet dataset (Jiao et al. 2020). In (Geirhos et al. 2018), it is found that the network trained directly on Imagenet actually relies on texture for classification.

- In terms of homogeneous transfer learning

In (Akram et al. 2020), a base model is firstly pre-trained from electroluminescence images dataset of photovoltaic cells and then fine-tuned on infrared images dataset. In (Gonzalez-Val et al. 2020), a large data set labeled with laser power and process speed is firstly constructed, and then trained ConvLBM in this dataset. The network can capture the dynamic changes of the laser process from the original image, replacing manual feature extraction. After the first stage of training is completed, new data is used

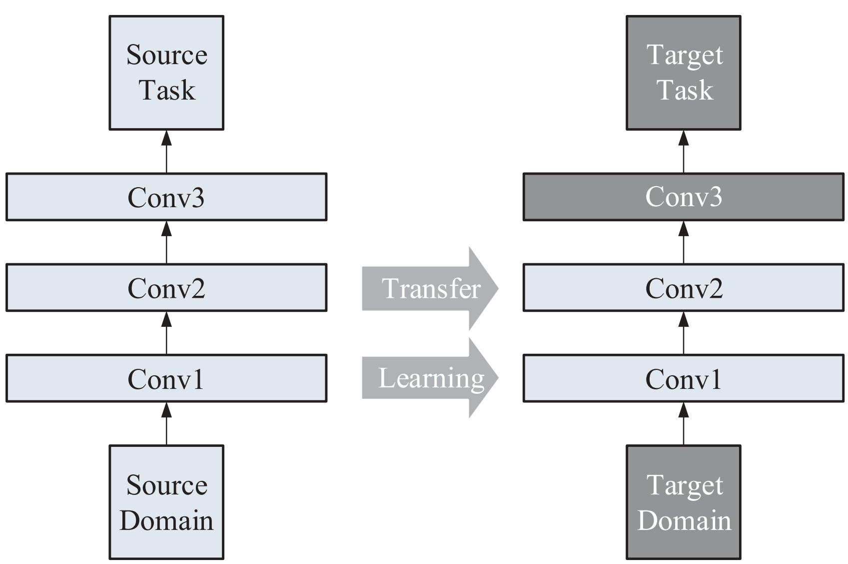

Figure 11. The connotation of transfer learning.

to train the existing ConvLBM model. This includes using images from the laser metal deposition process to train ConvLBM for estimating dilution, an indicator that could only be measured by destructive testing; and using laser welding images to train ConvLBM to detect defects in manufactured parts.

According to the connotation of transfer learning, the closer the source domain is to the target domain, the better the industrial application effect. In (G. Z. Fu et al. 2019), they believe that the shallow features of common images are not consistent with the shallow texture features on the surface of the steel plate. Therefore, in the fine-tuning stage, a larger learning rate is adopted for the features transferred from Imagenet. In addition, since pre-trained networks are used in transfer learning, the model architecture is limited. In other words, the convolutional layer cannot be removed in the pre-trained network optionally. When we import a pre-trained model, the network structure should be the same as the pre-trained network structure, and then train for specific scenarios and tasks.

3.7. Prior feature

The process of DL is to extract the feature patterns of the original image layer by layer. The basic features such as edges and colors are extracted at the bottom of the network, and some deep abstract features are extracted at the top of the network. In the case of insufficient information in sample space, it is impossible to extract reliable features effectively. Therefore, the prior feature is to artificially add some feature information to the sample space to simplify the learning difficulty of the network. The sources of priori feature mainly include manual features and automatic features based on deep learning. In terms of participation methods, the priori feature mainly participate in the two stages of feature extraction and feature classification.

As shown in Figure 12, there are three main ideas for introducing priori feature in the field of industrial images. The first is to artificially depth labeling the image (Figure 12-a). The idea is to carry out the tasks related to the target through the depth annotation of the image, and then take the result of the related task as the prior feature of the target task. The second is to use artificially extracted features as the input of CNN (Figure 12-b1). The third is that handcrafted features directly participate in the feature classification process (Figure 12-b2).

- In terms of depth annotation

In (X. Fu et al. 2020), the position of fasteners is firstly regressed and then the target area is segmented, so as to improve the segmentation accuracy. In (Tabernik et al. 2020; Racki, Tomazevic, and Skocaj

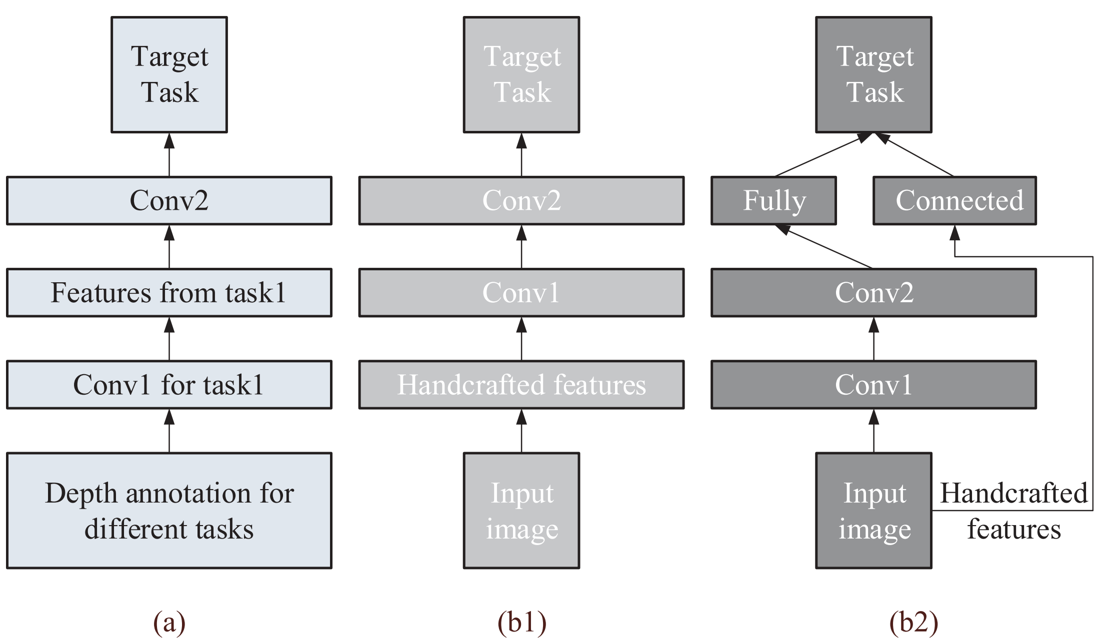

Figure 12. Two ways to introduce the priori features. a) Depth annotation; b) Handcrafted features.

2018; Tao et al. 2018), the proposed approach is formulated as a two-stage design. The first stage is to implement a segmentation network that performs a pixel-wise localization of the surface defect. The second stage, where binary-image classification is performed, includes an additional network that is built on top of the segmentation network and uses both the segmentation output as well as features of the segmentation net. In (Kang et al. 2019), a three-stage defect detection approach is proposed, which consists of key components localization, components segmentation and defect detection. In (Leary, Sawlani, and Mesbah 2020), CNN is first used to extract the features of scanning electron microscope image, and it is merged with the energy-dispersive X-ray data features of the wafer map in the classification stage. In (Kang et al. 2018), a Faster R-CNN network is adopted to localize the key catenary components, and then the classification score and anomaly score are determined from a deep multitask neural network.

- In terms of handcrafted features

In (Lwahori et al. 2018), feature points are generated by SIFT and the rectangular frame containing adjacent feature points is used as a candidate area. In this candidate area, a two-class SVM is first used to determine whether there is a defect. If it is defective by judging, a rectangular frame containing feature points is input to CNN for target detection. If the SVM finds that there is no defect, the region is still input to the CNN to be judged again to improve accuracy and robustness. In (Q. W. Liu et al. 2020), the candidate regions are generated by utilizing mean-shift clustering algorithm to the corners detected on binary images. Then, the CNN is used for the feature extraction and classification of candidate regions that possibly contain the aircraft, and the location of the aircraft is finally determined after further screening. H. Han et al. (2020) trained an RPN to get defect region proposals. Then they process the proposal regions to suitable size patches as the input of the modified U- net for defect segmentation. In (Ming et al. 2019), both HOG features and CNN-based automatically extracted features are used for defect detection of light guide plat. In (Y. X. Zhang et al. 2019), 351 manually extracted features are input to the CNN. In (Wang, Fu, and Gao 2019), the background is firstly removed by Hough transform and then the image is input to CNN. In (Kim et al. 2019), a vector output after the global average pooling layer is concatenated with handcrafted features and CNN features. The concatenated vector is used for classification through fully connected layer.

Although the introduction of prior feature by depth labeling can improve the accuracy of the target task, new tasks need to be processed in the outer functional layer, which leads to the increase of network size and operation time. The introduction of prior feature based on feature descriptors will result in the loss of the end-to-end efficient modeling capability of DL.

3.8. Summary

This chapter reviews the common methods and strategies of DLBIIPR from the perspective of how to enrich the sample space. From the current research perspective and content, this research is closer to the industrial image aspect. Its essence is to add additional information to the limited sample space to make the model have a stronger induction bias. The three methods of cross-validation, semi-supervised learning, and data augmentation are to enrich the sample space from the perspective of enriching the sample size. Moreover, cross validation is the full use of limited samples, semi-supervised learning is the use of unlabeled samples, and data augmentation directly generates new data. The four methods of multi-scale learning, attention mechanism, transfer learning, and priori feature are to enrich the sample space from the perspective of enriching the sample features. Moreover, multi-scale learning is to enrich the representation of features. The attention mechanism is to strengthen the useful features and weakening the useless features to enrich the sample space. Transfer learning is to introduce features from other datasets, and the prior feature is to enrich the sample features through human participation. Therefore, according to the evolution of data – feature – knowledge, this paper argues that in future research work it is possible to imitate human's cognitive process of images and add human's cognitive information to the sample space. Semantic information and temporal information are important cognitive sources. In (Yu et al. 2019), the features extracted by the pre-trained model are given the semantics of the basic features. In (Song, Song, and Yan 2020), they propose to integrate three types of high-level prior constraints (background, object, and midlevel feature constraint) to further boost the performance of the saliency detection. In (J. J. Zhang et al. 2018; L. T. Chen et al. 2020), a long short-term memory network is introduced on the problem of remaining life prediction to extract features of the time dimension.

4. Limiting the hypothesis space

The complexity of the industrial image itself makes the DLMs require a deep network to extract effective image patterns. This problem will not only increase the amount of calculation, but also easily cause overfitting problems. As a result, the feasibility and generalization ability of DLMs in practical applications will be reduced. In order to limit the hypothesis space, the current research ideas are mainly to limit the number of parameters and the range of fluctuation.

4.1. Parameter regularization

Parameter regularization is a common method to restrict the hypothesis space. Its essence is to limit the distribution of weight parameters, so as to reduce the hypothesis space. The formal description of parameter regularization is: Loss = L0 + $a * f$ (w), a is the coefficient of the regular term and f (w) is the parameter regularization term. a is a hyperparameter that needs to be set manually. L1-norm regularization introduces a priori knowledge that conforms to the Laplace distribution to the weight parameters. L2-norm regularization introduces a priori knowledge that conforms to the Gaussian distribution to the weight parameters. As shown in Figure 13, the effect of L1-norm regularization is to make the weight parameters sparse, and the effect of L2-norm regularization is to make the weight parameters smaller and smoother. So, L1-norm regularization tends to limit the solution space by reducing the number of parameters, while L2-norm regularization tends to limit the solution space by restricting the range of values of the parameters. Therefore, we distinguished the two norm regularization methods in the introduction. However, in order to facilitate the overview, we put the two methods together in this section.

The L1-norm regularization (Y. He et al. 2019bb; L. Guo et al. 2019; Shilon et al. 2019) and L2-norm regularization (T. He et al. 2020; Akram et al. 2020; Kang et al. 2018; Wei et al. 2019) are commonly used in the industrial field. The authors (Guo et al. 2019a) empirically set a to 2^{-30}. The authors (Wei et al. 2019) added L2-norm regularization to the fully connected layer and set the coefficient a to 0.0005.

Although L1 and L2 parameter regularization has been widely used in DLBIIPR, there are still the following shortcomings. Firstly, the coefficient of the regularization term is a hyperparameter, which needs to be carefully optimized in the application process. Secondly, the L1-norm regularization and L2-norm regularization are the methods for traditional neural networks. CNN is a commonly used architecture in DLBIIPR, while the traditional L1 and L2 regularization do not consider the characteristics of CNN. Therefore, it is necessary to take into account the architectural features of CNN to design the corresponding parameter regularization methods. In (Yang et al. 2019), it is considered that many features extracted by convolution kernels are similar and unnecessary, so sparse and entropy similarity constraints are added to convolution kernels. In (T. Y. Liu et al. 2020), the feature of local association of weight parameters in CNN was considered. In order to maximize the ability of convolutional kernels to express molten pool features, a coarse grained regularization method was proposed that convolutional kernels should be trained in the direction of difference maximization.

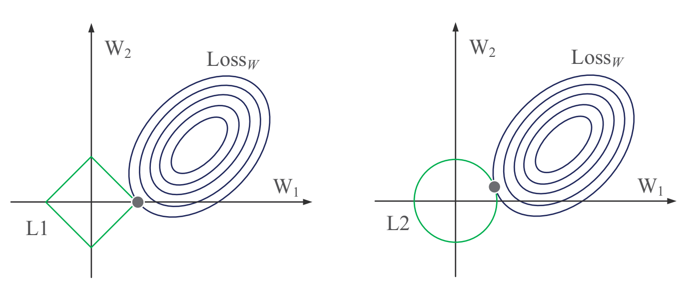

Figure 13. Parameter regularization.

4.2. Dropout

Dropout is a stochastic technology used in neural networks. Dropout methods commonly used in deep learning include standard Dropout, CNN-oriented Dropout, and RNN-oriented Dropout. The dropout method not only accelerates the training process of the network, but also reduces the hypothesis space by discarding the neurons.

As shown in Figure 14, standard dropout is to stop the activation value of a neuron with a probability p during the forward propagation of the neural network. Then the loss is propagated back according to the network, and the neurons that do not stop working are updated. This method can greatly reduce the hypothesis space. In this case, the updating of weights no longer depends on the joint action of hidden nodes with fixed relationships, which prevents certain features from having effects only under other specific features. Since different neurons have been chosen to be discarded in each iteration, the resulting model is equivalent to the average of multiple simple models. Therefore, the network can attain more robust features.

Standard dropout is still widely used in current industrial applications. The most used in related work is to standard dropout with a probability of 0.5 (T. He et al. 2020; Meijer et al. 2019; Z. H. Zhang et al. 2020). In addition, dropout is performed with the probability of 0.3 (Kang et al. 2018), 0.6 (Xie et al. 2019) and 0.7 (Kang et al. 2019).

However, the discard rate is a hyperparameter, and the selection of this hyperparameter is currently lacking in-depth study in the DLBIIPR study. What's more, this method is not designed according to the characteristics of CNN. Therefore, in future industrial applications, more consideration should be given to using the Dropout method that combines the characteristics of CNN. There are five main dropout methods for CNN: drop by block (Ghiasi, Lin, and Le 2018), drop by feature map (Tompson et al. 2015), drop blocks or feature maps with high activation (Park and Kwak 2016), apply random square occlusion for input (Devries and Taylor 2017), and drop elements in pooling blocks (Wu and Gu 2015).

4.3. Efficient modules

- Batch Normalization

Since the essence of the DL process is to learn the distribution of data, once the distribution of training data is different from testing data, the generalization ability of the network will be greatly reduced. In addition, once the distribution of each batch of training data is different, the network will need to learn a different distribution in each iteration, which will increase the parameter search space and reduce the training speed and generalization ability of the network. Therefore, in order to limit the size of the hypothesis space and increase the speed of parameter optimization, a batch normalization layer (BN) will be inserted before inputting the next layer (Loffe and Szegedy 2015). Although the core idea of BN is similar to image normalization, the BN layer is an intermediate layer with learnable parameters. The BN layer makes the distribution of input data of each layer in the network relatively stable, which can limit the fluctuation of the gradient. Therefore, the optimization space can be smoother, thereby improving the generalization ability of the DLMs. This method has been widely used in the industrial field (Akram et al.

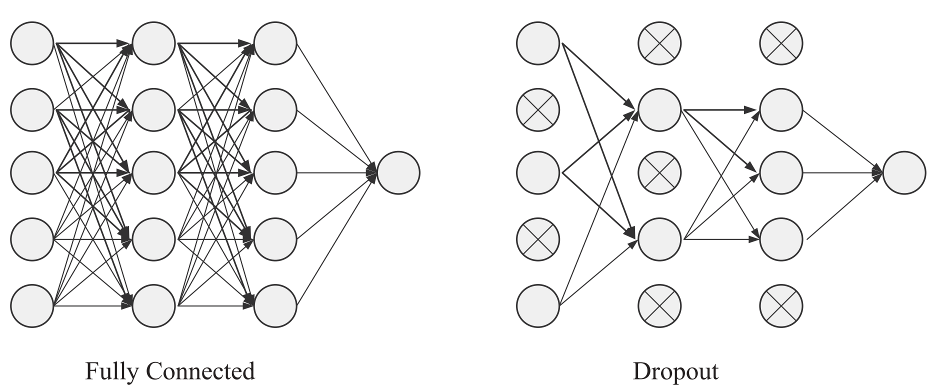

Figure 14. Standard dropout.

2020; Gonzalez-Val et al. 2020; Xu et al. 2020; Z. P. Guo et al. 2019; Q. K. Liang et al. 2019; J. X. Wu et al. 2020; Zhang et al. 2020b; Wang et al. 2020b).

- Depthwise separable Convolution

As shown in Figure 15, the depthwise separable convolution decompose standard convolution into a depthwise convolution and a pointwise convolution (Howard et al. 2017). This method implies a priori information that the spatial location of data is highly correlated but different channels are independent of each other. Compared with the standard convolution, the depthwise separable convolution will greatly reduce the parameters, so the size of the hypothesis space can be reduced while keeping the network accuracy substantially unchanged. The depthwise separable convolution is also widely used in the industrial field (Z. P. Guo et al. 2019; X. Y. Chen et al. 2020; Xia, Cao, and Wang 2019; Hu and Wang 2019; J. J. Wang et al. 2019).

Since IIPR is an application of DL, the development of DL itself is faster than the application process. Therefore, the future DLBIIPR needs to pay more attention to the latest DL technology to improve the application effect. For example, Zhang et al. (2017) added information exchange between channels. Sabour, Frosst, and Hinton (2017) proposed to use vectors instead of scalars to represent feature points, and Hunag et al. (2017) proposed a method for

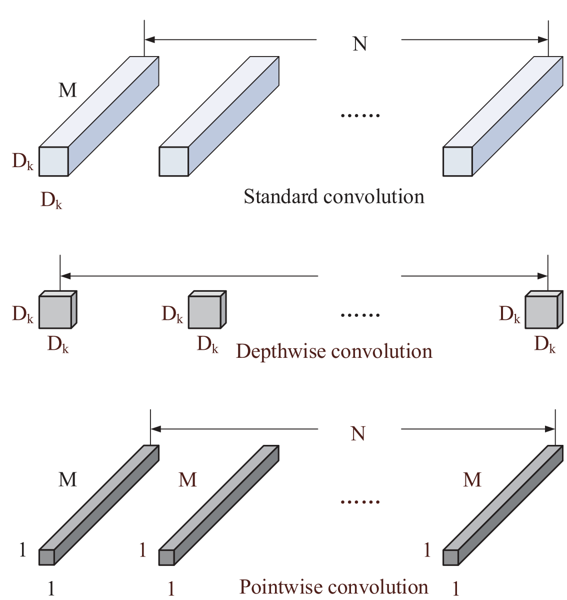

Figure 15. Depthwise separable convolution.

automatically learning input feature grouping during training.

4.4. Image normalization

Each pixel in each image can be seen as a feature point, and the range of fluctuation in different pixel points in an image tends to vary widely. The normalization of the input image is to make the feature scale of each pixel point within a range, so that the back propagation error of each feature point will not be too different. Therefore, the hypothesis space can be compressed so that the DLMs can more easily converge to the optimal solution. This method plays a similar function to L2-norm regularization.

The authors (Leary, Sawlani, and Mesbah 2020) scaled the grayscale value from [0, 255] to [0,1] and used the Relu activation function. In (Zhao et al. 2020; Akram et al. 2020), the authors also used Relu as the activation function, and the pixel range is compressed from [0, 255] to [0, 1]. In (Zhang et al. 2019c), the authors manually extracted 351 feature values, and then the sample values of each feature are normalized to [0, 1] in order to improve the accuracy and robustness of the DLMs.

For DL, although image normalization can make the sample space regular, different activation functions respond differently to different numerical intervals, so different normalization intervals should be selected based on different activation functions in practice. However, in the current DLBIIPR research, researchers lack the optimal choice of activation function after adopting the image normalization method.

4.5. Early stopping

For many optimization algorithms used for DL training (such as gradient descent), the loss function is a decreasing function about the number of epochs. However, the error on the validation set usually decreases firstly and then gradually increases due to overfitting of the DLM. As shown in Figure 16, the basic idea of early stopping method is to stop training when the model's performance on the validation set starts to decline (can be customized). In this way, a model with good generalization performance can be obtained. There are three commonly used stopping rules in DL. The first is to stop training when the generalization loss exceeds a certain threshold.

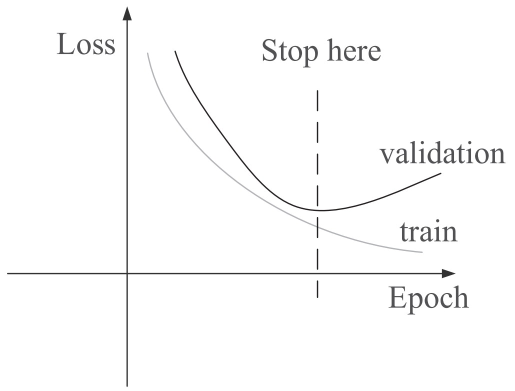

Figure 16. Early stopping.

The second kind completely depends on the change of generalization loss, that is, the training is stopped when the generalization loss increases in consecutive *k* epochs. The third type is determined by the generalization loss and its changing speed. From the perspective of restricting the hypothesis space, it can be considered that this method limits the parameter space of the optimization process to the small neighborhood of the initial parameters.

In (Meijer et al. 2019), the network stopped training when the AUROC on the validation fold had not increased for 25 consecutive assessments, or when the AUROC on the validation fold decreased by>1%, with a minimum of 1000 batches processed. In (Imoto et al. 2018), they reduced the runtime of training with hundreds of thousands of images from 40 h to 4 h, increasing the speed tenfold by combining use of early stopping and GPU parallel computing.

The disadvantage of early stopping is that it does not take different methods to solve these two problems of optimizing the loss function and reducing the validation loss. Instead, it uses one method to solve these two problems at the same time. This may face the problem that the training loss is not low enough.

4.6. Multi-task learning

The main goal of multi-task learning (MTL) is to improve the generalization ability by utilizing the specific domain information implicit in the training signals of multiple related tasks. The learned model emphasizes the common representation among multi-tasks. As shown in Figure 17, the common MTL models in DL have hard parameter sharing mechanism and soft parameter sharing mechanism. The shared representation in related tasks makes the DLMs have a stronger inductive bias for certain hypotheses, so it can play an important role in limiting the hypothesis space, thereby improving the generalization ability of the DLMs.

In (D. He et al. 2019aa), the defect classification task and the defect image reconstruction task were jointly trained and the parameters between the two tasks were restricted by soft sharing. In (Xu et al. 2019), the framework consisted of three parts: shared layers and two sub-networks, as well as a location network (LNet) and a decision network (DNet). The shared layer is used to extract the feature representation of the defect and serves as the input to the LNet. The LNet is used to determine the probability that a defect exists at the corresponding location in the original image. The DNet is used to predict the probability of

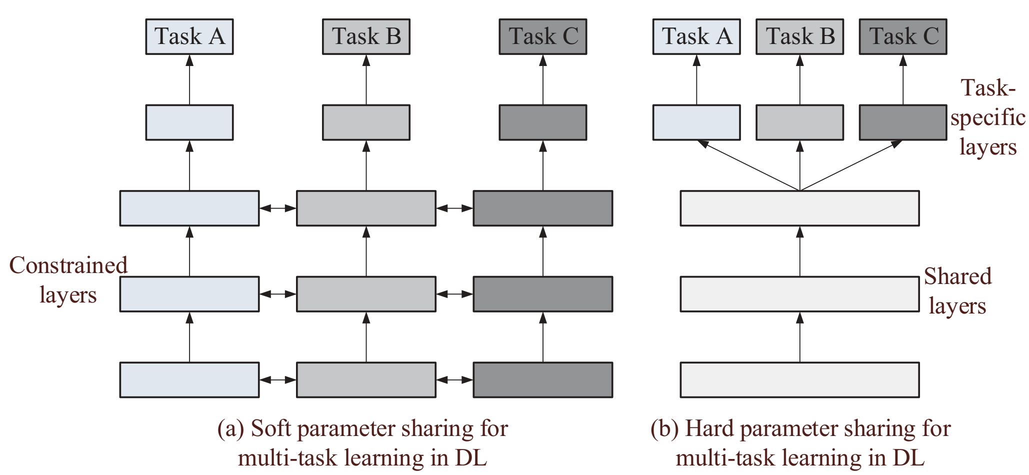

Figure 17. Multi-task learning.

defects being present at specific locations. In (Kang et al. 2018), a deep material classifier and a deep denoising autoencoder were integrated into a deep multitask neural network architecture, needing no defective samples for training to accomplish simultaneous segmentation and defect detection. In (Han et al. 2020a; Du et al. 2019), the two tasks of defect classification and position regression were jointly trained. In (Shen et al. 2020), the method had three main tasks: Faster RCNN as base network for defect object detection, a segmentation network for aluminum area segment and a multi-label classification network for defect labeling.

The hard parameter sharing mechanism is widely used in engineering applications. The first thing noted in engineering applications is that this way of sharing knowledge between tasks is useful when the tasks are very similar. However, the model performance will be deteriorated when this assumption is violated. How to balance the losses from multi-tasks is another issue that needs pay attention to in applications. In (Kendall, Gal, and Cipolla 2017), they proposed a method to transform different loss into a uniform scale for fusion.

4.7. Summary

This chapter reviews the common methods and strategies of DLBIIPR from the perspective of how to limit the hypothesis space. The L1 parameter regularization, dropout, and efficient modules reduce the size of the hypothesis space by limiting the number of parameters. Among them, L1 regularization is a method for parameters, dropout is a method for neurons, and efficient modules are the methods for network architectures. The image normalization, L2 parameter regularization, early stopping and multi-task learning are to reduce the size of the hypothesis space from the perspective of restricting the parameter fluctuation range. Among them, image normalization is a method for the input, L2 regularization is a method for the parameters, early stopping is a method for the training process and multi-task learning is a method for the network architecture. From the current research perspective and content, this research is closer to the deep learning side. The key technologies involved in this idea mostly come from the development of deep learning and convolutional neural networks. The essence of related technologies is to restrict the hypothesis space through related prior knowledge. In the future DLBIIPR, the latest DL methods can be combined with the specific characteristics of industrial images to limit the model hypothesis space while satisfying the feature representation.

5. Future challenges and possible issue

5.1. Requirement analysis

The methods for enriching the sample space and limiting the hypothesis space have been reviewed previously. Through the joint use of related methods, the feasibility and generalization ability of DLBIIPR have been greatly improved. Although DLMs can perform well on the testing dataset, they still face the following problems in practical applications. Firstly, due to biases in the dataset (e.g. not meeting the assumption of IID), we cannot guarantee the reliability of DLMs by the testing accuracy (Zhang, Wang, and Zhu 2017), and such a single evaluation metric is insufficient for a complete description of many realistic tasks (Doshi-Velez and Kim 2017). Secondly, the accuracy derived from DL methods is a concept in the sense of probability and statistics. Thirdly, over time, various factors in the environment can cause the conceptual drift of DLMs and the performance of DL will also keep changing. In industry, the main focus of DL should be more on solving complex industrial problems in an 'applied' way rather than theoretically applying DL methods to the correct data. However, in current industrial applications, the DLMs are like a black box for the front-line workers. The lack of explainability severely limits the wide application of DL methods in safety-sensitive tasks such as automated driving, aerospace. In addition, unexplainability also means insecure, that is, it cannot effectively resist attacks on the DLMs. Thus, the lack of explainability has become one of the main obstacles to the further development and application of DL in industrial image (M. Ibrahim et al. 2019).

XDL is used to explain the basis of DLMs' decisions to ensure that the reliability of DLMs is guaranteed because of explainability (Ribeiro, Singh, and Guestrin 2016). The reliability of DLMs can be evaluated and validated at a fine-grained by using explainable models and related explainable methods, which help us to understand more clearly why the DLMs make certain decisions and possible problems. Over time, XDL helps build trust between users and DLMs and eliminates security risks in real-world applications of the DLMs. In general, the more complex the DLMs are, the more expressive they will be, but the explainability of the DLMs will become weaker and weaker. Moreover, due to the limitation of the characteristics of industrial image, the feasibility and generalization ability of DLMs will become weaker and weaker. The explainability research of DL is oriented to the interaction between human and DLMs. This interactive process will introduce human prior information, so XDL can help academia optimize DLMs to improve the feasibility and generalization ability.

5.2. Status analysis

'Interpretable' and 'Explainable' are two expressions that appear frequently in the literature. Interpretable AI is the process of making an inherently transparent model (white box) understandable, while Explainable AI is the process by which the researcher provides an explanation for the behavior of the black box model that people can understand (Lecue 2019). In the case of DL, it is obviously more suitable for the expression of 'Explainable'. The definition of explainability has not been completely consistent, and there are the following definitions related to XDL. Doshi-Velez and Kim (2017) considered explainable capacity as the ability to interpret or express in a way that human can understand. Biran and Cotton (2017) considered that a key component of an artificially intelligent system is the ability to explain the decisions, recommendations, predictions or actions made by it and the process through which they are made. In the context of ML systems, Doshi-Velez and Kim (2018) added an emphasis on providing explanation to human, that is, to explain or to present in understandable terms to a man. Kim and Doshi-Velez (2017) considered that interpretation is the process of giving explanations to human. What's more, interpretability does not exist for an explicit goal (such as an objective function to be optimized), but to ensure that some aspects are safeguarded by interpretability itself. In (Guidotti et al. 2018), explainability was considered to be the interface between frontline users and the DLMs, which is an accurate proxy for the decision model and understandable to human. One of the key points in the above definition is that XDL needs to provide a way that human can understand.

Academics divide explainability methods into ante-explainable (ante-hoc) and post-explainable (post-hoc) (Lipton 2018; Došilović, Brcic, and Hlupic 2018). Ante-hoc refers to the construction of white box models that are inherently explainable, such as naive Bayes, linear regression, decision trees, and rule-based models. Post-hoc refers to the interpretation of the trained DLMs by developing explainability techniques. In the case of DL, it is more applicable to post-hoc methods. According to the purpose and object of explainability, post-hoc can be divided into global explaination and local explaination. Global explaination aims to help people understand the complex logic behind the model as a whole, as well as the internal working mechanism, such as how the model learns, what the model learns from the training data, how the model makes decisions, etc., which requires us to be able to represent a trained complex learning model in a human-understandable way. Typical globally explainable methods include rule extraction (L. M. Fu 1994), model distillation (Hinton, Vinyals, and Dean 2015), and activation maximization (Simonyan, Vedaldi, and Zisserman 2013). Local explaination aims to help people understand the decision process and decision basis of DLMs for each specific sample. In contrast to global explaination, local explaination is sample-oriented and can usually be achieved by analyzing the contribution of each feature dimension of the input sample to the final decision outcome of the model. In practice, due to the opacity and complexity of the model and the multiplicity of application scenarios, it is usually more difficult to provide global explanation than local explanation, and thus local explanation studies are more widespread and more common. Typical locally explaination methods include sensitivity analysis (Sung 1998), local approximation (Ribeiro, Singh, and Guestrin 2016, 2018), back propagation (Springenberg et al. 2014), feature inversion (M. N. Du et al. 2018), and class activation mapping (Zhou et al. 2016).

5.3. Case study

The explainable methods are critical for developers to understand the basis for decisions and optimize DLMs. An example of radiographic image inspection of a weld containing blowhole and lack of penetration defects is presented. Figure 18 shows the saliency map based on back propagation and the feature map based on feature inversion. The core idea of the back propagation-based explainable method is to use the back propagation mechanism to propagate the decision signal from the output layer to the input layer of the DLM layer by layer to derive the salient regions in the input sample. The feature inversion method can make full use of the middle layer information of the DLM to provide an explanation for the decision-making behavior of the DLM.

It can be seen from the above figure that when the trained DLM makes a decision on the defect category, the back propagation method can infer important regions in the input image, and the feature inversion method can accurately locate the important features used for model decision in the input instance. Although the considerations of each approach are different, all of them can help developers understand the decision-making process of DLM. Figure 18-b1 and 18-c1 show an explanation of the decision behavior of a DLM with mediocre performance. It can be seen that the DLM is susceptible to noise and pays insufficient attention to the lack of penetration regions where the gray scale does not vary significantly. After optimizing the DLM by adding attention mechanism, it can be seen from Figure 18-b2 and 18-c2 that the DLM has basically placed the decision basis on the two defect areas of blowhole and lack of penetration.

6. Conclusions

Enriching the information in the sample space and limiting the size of the hypothetical space are two basic ideas for improving the feasibility and generalization performance of DLMs. Based on these two ideas, this paper reviews the specific methods for

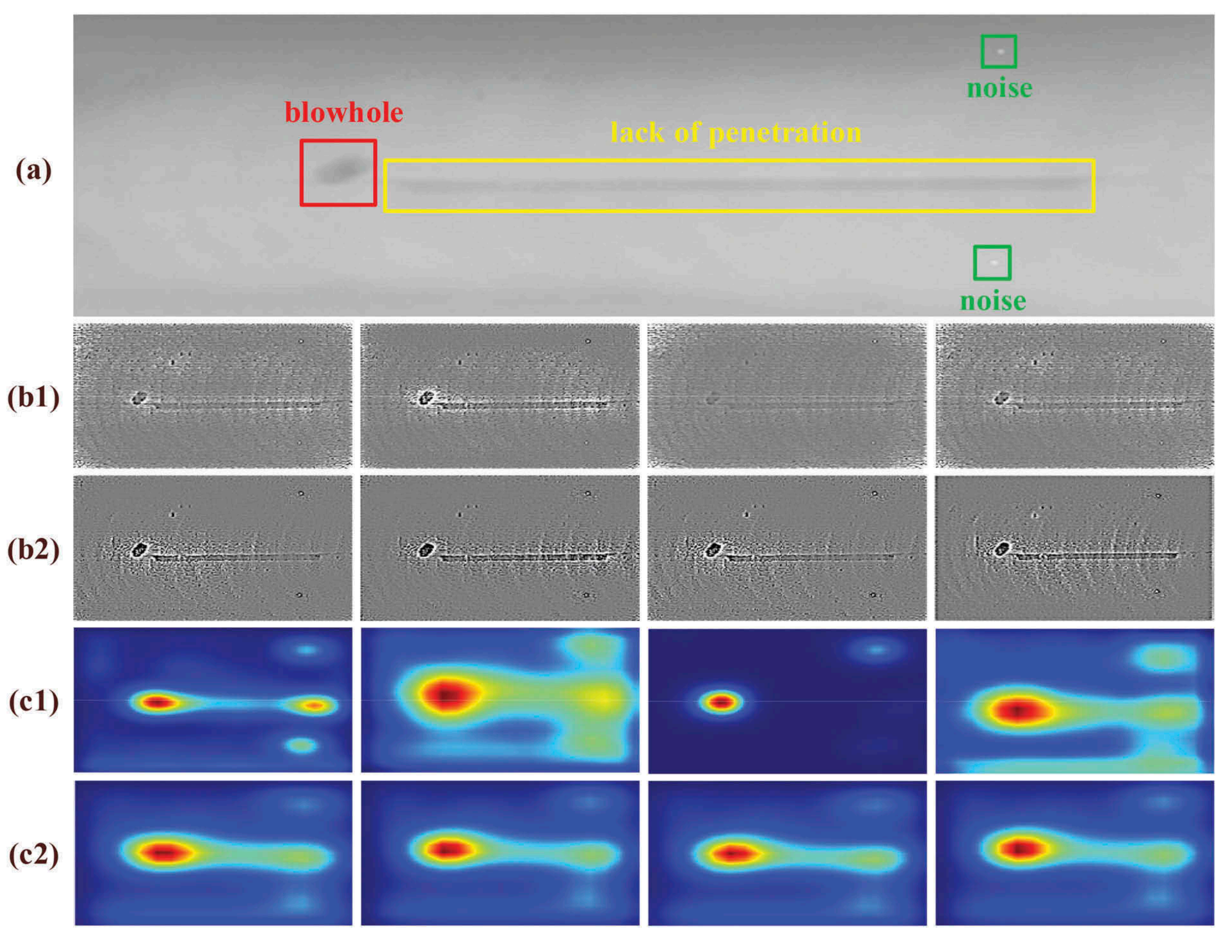

Figure 18. The back propagation-based and feature inversion-based explainable. a) Original image; b) Feature inversion-based; c) Back propagation-based.

improving the feasibility and generalization ability of DLBIIPR.

In terms of enriching the information in the sample space, related research is closer to the industrial image side. Its essence is to add additional information to the limited sample space so that the DLMs will have a clear induction bias. In terms of limiting the hypothesis space, related research is closer to the DL side. Its essence is to constrain the hypothesis space through related prior knowledge so that the DLMs may be simple on the basis of satisfying the feature representation (the Occam's Razor principle). Its key technologies mostly come from the development of DL, especially convolutional neural networks. In future research, this article argues that the characteristics of industrial image needs to be analyzed first. Then we need to add the corresponding prior information according to the requirements of the application scenario to improve the algorithm's inductive bias. What's more, explainability not only assists developers in evaluating and improving DLMs but also improves the reliability of DLMs. In that regard, this paper can guide more researchers to address the current and future directions of the DL research and application in industrial image.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Natural Science Foundation of China [51905091]; The Fundamental Research Funds for the Central Universities and Graduate Student Innovation Fund of Donghua University [CUSF-DH-D-2020053].

ORCID

Jinsong Bao http://orcid.org/0000-0003-1999-1003

References

Ahmad, J., A. Akula, R. Mulaveesala, and H. K. Sardana. 2019. "An Independent Component Analysis Based Approach for Frequency Modulated Thermal Wave Imaging for Subsurface Defect Detection in Steel Sample." *Infrared Physics & Technology* 98: 45–54. doi:10.1016/j.infrared.2019.02.006.

Akram, M. W., G. Q. Li, Y. Jin, X. Chen, C. G. Zhu, and A. Ahmad. 2020. "Automatic Detection of Photovoltaic Module Defects in Infrared Images with Isolated and Develop-model Transfer Deep Learning." *Solar Energy* 198: 175–186. doi:10.1016/j.solener.2020.01.055.

Alex, K., I. Sutskever, and G. E. Hinton. 2012. "ImageNet Classification with Deep Convolutional Neural Networks." In: Proceedings of the 25th International Conference on Neural Information Processing Systems (NIPS 2012), Lake Tahoe, December 3–6. doi: 10.1145/3065386.

Alexopoulos, K., N. Nikolakis, and G. Chryssolouris. 2020. "Digital Twin-driven Supervised Machine Learning for the Development of Artificial Intelligence Applications in Manufacturing." *International Journal of Computer Integrated Manufacturing* 33 (5): 429–439. doi:10.1080/0951192X.2020.1747642.

Arlot, S., and A. Celisse. 2010. "A Survey of Cross-validation Procedures for Model Selection." *Statistics Surveys* 4: 40–79. doi:10.1214/09-SS054.

Bacioiu, D., G. Melton, M. Papaelias, and R. Shaw. 2019. "Automated Defect Classification of SS304 TIG Welding Process Using Visible Spectrum Camera and Machine Learning." *NDT & E International* 107: 102139. doi:10.1016/j.ndteint.2019.102139.

Biran, O., and C. Cotton. 2017. "Explanation and Justification in Machine Learning: A Survey." In: Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence (IJCAI 2017), Melbourne, August 19–25. http://www.cs.columbia.edu/~orb/papers/xai_survey_paper_2017.pdf.

Buerhop-Lutz, C., S. Deitsch, A. Maier, F. Gallwitz, S. Berger, B. Doll, J. Hauch, C. Camus, and C. J. Brabec. 2018. "A Benchmark for Visual Identification of Defective Solar Cells in Electroluminescence Imagery." In: 35th European photovoltaic solar energy conference and exhibition (EU PVSEC 2018), Brussels, September 24–28. doi: 10.4229/35thEUPVSEC20182018-5CV.3.15.

Chen, L. T., G. H. Xu, S. Zhang, W. Q. Yan, and Q. Q. Wu. 2020a. "Health Indicator Construction of Machinery Based on End-to-end Trainable Convolution Recurrent Neural Networks." *Journal of Manufacturing Systems* 54: 1–11. doi:10.1016/j.jmsy.2019.11.008.

Chen, X. Y., J. Y. Chen, X. G. Han, C. D. Zhao, D. Z. Zhang, K. F. Zhu, and Y. J. Su. 2020b. "A Light-weighted CNN Model for Wafer Structural Defect Detection." *IEEE Access* 8: 24006–24018. doi:10.1109/ACCESS.2020.2970461.

Chen, Y. P., H. Q. Fang, B. Xu, Z. C. Yan, Y. Kalantidis, M. Rohrbach, S. C. Yan, and J. S. Feng. 2019. "Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution." arXiv preprint arXiv: 1904.05049v1, https://arxiv.org/abs/1904.05049v1.

Cheon, S., H. Lee, C. O. Kim, and S. H. Lee. 2019. "Convolutional Neural Network for Wafer Surface Defect Classification and the Detection of Unknown Defect Class." *IEEE Transactions on Semiconductor Manufacturing* 32 (2): 163–170. doi:10.1109/TSM.2019.2902657.

Cui, D., G. Q. Wang, Y. P. Lu, and K. K. Sun. 2020. "Reliability Design and Optimization of the Planetary Gear by a GA

Based on the DEM and Kriging Model." *Reliability Engineering & System Safety* 203: 107074. doi:10.1016/j.ress.2020.107074.

Danelljan, M., G. Bhat, F. S. Khan, and M. Felsberg. 2017. "ECO: Efficient Convolution Operators for Tracking." In: IEEE conference on computer vision and pattern recognition (CVPR 2017), Honolulu, July 21–26. doi: 10.1109/CVPR.2017.733.

Devries, T., and G. W. Taylor. 2017. "Improved Regularization of Convolutional Neural Networks with Cutout." arXiv preprint arXiv: 1708.04552, https://arxiv.org/abs/1708.04552.

Donati, L., E. Iotti, G. Mordonini, and A. Prati. 2019. "Fashion Product Classification through Deep Learning and Computer Vision." *Applied Science* 9 (7): 1385. doi:10.3390/app9071385.

Doshi-Velez, F., and B. Kim. 2017. "Towards a Rigorous Science of Interpretable Machine Learning." arXiv preprint arXiv: 1702.08608, https://arxiv.org/abs/1702.08608.

Doshi-Velez, F., and B. Kim. 2018. *Explainable and Interpretable Models in Computer Vision and Machine Learning: Considerations for Evaluation and Generalization in Interpretable Machine Learning*. Cham: Springer.

Došilović, F. K., M. Brcic, and N. Hlupic. 2018 "Explainable Artificial Intelligence: A Survey." In: 41st international convention on information and communication technology, electronics and microelectronics (MIPRO 2018), Opatija, May 21–25. doi: 10.23919/MIPRO.2018.8400040.

Du, M. N., N. H. Liu, Q. Q. Song, and X. Hu. 2018. "Towards Explanation of Dnn-based Prediction with Guided Feature Inversion." In: 24th ACM SIGKDD international conference on knowledge discovery & data mining (KDD 2018), Lonton, August 19–23. doi: 10.1145/3219819.3220099.

Du, W. Z., H. Y. Shen, J. Z. Fu, G. Zhang, and Q. He. 2019. "Approaches for Improvement of the X-ray Image Defect Detection of Utomobile Casting Aluminum Parts Based on Deep Learning." *NDT & E International* 107: 102144. doi:10.1016/j.ndteint.2019.102144.

Du, W. Z., H. Y. Shen, J. Z. Fu, G. Zhang, X. K. Shi, and Q. He. 2020. "Automated Detection of Defects with Low Semantic Information in X-ray Images Based on Deep Learning." *Journal of Intelligent Manufacturing*. doi:10.1007/s10845-020-01566-1.

Dung, C. V., H. Sekiya, S. Hirano, T. Okatani, and C. Miki. 2019. "A Vision-based Method for Crack Detection in Gusset Plate Welded Joints of Steel Bridges Using Deep Convolutional Neural Networks." *Automation in Construction* 102: 217–229. doi:10.1016/j.autcon.2019.02.013.

Fabirc Defect Dataset of Aliyun Tianchi Competition (11. 26, 2019). https://tianchi.aliyun.com/competition/entrance/231666/information.

Fan, H., P. C. Cosman, Y. Jou, and B. Li. 2018. "High-Speed Railway Fastener Detection Based on a Line Local Binary Pattern." *IEEE Signal Processing Letters* 25 (6): 788–792. doi:10.1109/LSP.2018.2825947.

Feng, W., H. D. Liu, D. X. Zhao, and X. J. Xu. 2020. "Research on Defect Detection Method for High-reflective-metal Surface Based on High Dynamic Range Imaging." *Optik* 206: 164349. doi:10.1016/j.ijleo.2020.164349.

Fu, G. Z., P. Z. Sun, W. B. Zhu, J. X. Yang, Y. L. Cao, M. Y. Yang, and Y. P. Cao. 2019. "A Deep-learning-based Approach for Fast and Robust Steel Surface Defects Classification." *Optics and Lasers in Engineering* 121: 397–405. doi:10.1016/j.optlaseng.2019.05.005.

Fu, L. M. 1994. "Rule Generation From Neural Networks." *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 24 (8): 1114–1124. doi:10.1109/21.299696.

Fu, X., K. L. Li, J. Liu, K. Q. Li, Z. Zeng, and C. Chen. 2020. "A Two-stage Attention Aware Method for Train Bearing Shed Oil Inspection Based on Convolutional Neural Networks." *Neurocomputing* 380: 212–224. doi:10.1016/j.neucom.2019.11.002.

Gan, J. R., Q. Y. Li, J. Z. Wang, and H. M. Yu. 2017. "A Hierarchical Extractor-based Visual Rail Surface Inspection System." *IEEE Sensors Journal* 17 (23): 7935–7944. doi:10.1109/JSEN.2017.2761858.

Gao, Y. P., L. Gao, X. Y. Li, and X. V. Wang. 2020b. "A Multi-Level Information Fusion-based Deep Leaning Method for Vision-based Defect Recognition." *IEEE Transactions on Instrumentation and Measurement* 69 (7): 3980–3991. doi:10.1109/TIM.2019.2947800.

Gao, Y. P., L. Gao, X. Y. Li, and X. G. Yan. 2020a. "A Semi-supervised Convolutional Neural Network-based Method for Steel Surface Defect Recognition." *Robotics and Computer-Integrated Manufacturing* 61: 101825. doi:10.1016/j.rcim.2019.101825.

Geirhos, R., P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel. 2018. "ImageNet-trained CNNs are Biased Towards Texture; Increasing Shape Bias Improves Accuracy and Robustness." arXiv preprint arXiv: 1811.12231, https://arxiv.org/abs/1811.12231.

Ghiasi, G., T. Y. Lin, and Q. V. Le. 2018. "DropBlock: A Regularization Method for Convolutional Networks." arXiv preprint arXiv: 1810.12890, https://arxiv.org/abs/1810.12890.

Girshick, R. 2015 "Fast R-CNN." In: IEEE international conference on computer vision (ICCV 2015), Santiago, December 7–13. doi: 10.1109/ICCV.2015.169.

Girshick, R., J. Donahue, T. Darrell, and J. Malik. 2014. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." In: IEEE conference on computer vision and pattern recognition (CVPR2014), Columbus, June 23–28. doi: 10.1109/CVPR.2014.81.

Gonzalez-Val, C., A. Pallas, V. Panadeiro, and A. Rodriguez. 2020. "A Convolutional Approach to Quality Monitoring for Laser Manufacturing." *Journal of Intelligent Manufacturing* 31 (3): 789–795. doi:10.1007/s10845-019-01495-8.

Goodfellow, I. J., J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. "Generative Adversarial Networks." arXiv preprint arXiv: 1406.2661, https://arxiv.org/abs/1406.2661.

Guan, S. Y., X. K. Wang, L. Hua, and L. Li. 2021. "Quantitative Ultrasonic Testing for Near-surface Defects of Large Ring Forgings Using Feature Extraction and GA-SVM." *Applied Acoustics* 173: 107714. doi:10.1016/j.apacoust.2020.107714.

Guidotti, R., A. Monreale, S. Ruggieri, F. Turini, D. Pedreschi, and F. Giannotti. 2018. "A Survey of Methods for Explaining Black Box Models." *ACM Computing Surveys* 51 (5): 1–42. doi:10.1145/3236009.

Guo, L., R. Z. Li, B. Jiang, and X. Shen. 2019a. "Automatic Crack Distress Classification from Concrete Surface Images Using a Novel Deep-width Network Architecture." *Neurocomputing* 397: 383–392. doi:10.1016/j.neucom.2019.08.107.

Guo, Z. P., Z. W. Wu, S. Liu, X. Ma, C. Y. Wang, D. J. Yan, and F. L. Niu. 2019b. "Defect Detection of Nuclear Fuel Assembly Based on Deep Neural Network." *Annals of Nuclear Energy* 137: 107078. doi:10.1016/j.anucene.2019.107078.

Han, H., C. Q. Gao, Y. Zhao, S. S. Liao, L. Tang, and X. D. Li. 2020a. "Polycrystalline Silicon Wafer Defect Segmentation Based on Deep Convolutional Neural Networks." *Pattern Recognition Letters* 130: 234–241. doi:10.1016/j.patrec.2018.12.013.

Han, Q., S. C. Wang, Y. Fang, L. Wang, X. Y. Du, H. L. Li, Q. X. He, and Q. B. Feng. 2020b. "A Rail Fastener Tightness Detection Approach Using Multi-source Visual Sensor." *Sensors* 20 (5): 1367. doi:10.3390/s20051367.

Hassanin, A. A., E. S. F. Abd, and E. B. Ghada. 2019. "A Real-time Approach for Automatic Defect Detection from PCBs Based on SURF Features and Morphological Operations." *Multimedia Tools and Applications* 78 (24): 1–21. doi:10.1007/s11042-019-08097-9.

He, D., K. Xu, and D. D. Wang. 2019a. "Design of Multi-scale Receptive Field Convolutional Neural Network for Surface Inspection of Hot Rolled Steels." *Image and Vision Computing* 89: 12–20. doi:10.1016/j.imavis.2019.06.008.

He, D., K. Xu, P. Zhou, and D. D. Zhou. 2019b. "Surface Defect Classification of Steels with a New Semi-supervised Learning Method." *Optics and Lasers in Engineering* 117: 40–48. doi:10.1016/j.optlaseng.2019.01.011.

He, K. M., X. Zhang, S. Q. Ren, and J. Sun. 2016. "Deep Residual Learning for Image Recognition." In: IEEE conference on computer vision and pattern recognition (CVPR 2016), Las Vegas, June 27–30. doi: 10.1109/CVPR.2016.90.

He, T., Y. Liu, Y. B. Yu, Q. Zhao, and Z. K. Hu. 2020. "Application of Deep Convolutional Neural Network on Feature Extraction and Detection of Wood Defects." *Measurement* 152: 107357. doi:10.1016/j.measurement.2019.107357.

He, Y., K. C. Song, H. W. Dong, and Y. H. Yan. 2019a. "Semi-supervised Defect Classification of Steel Surface Based on Multi-training and Generative Adversarial Network." *Optics and Lasers in Engineering* 122: 294–302. doi:10.1016/j.optlaseng.2019.06.020.

He, Y., K. C. Song, Q. G. Meng, and Y. H. Yan. 2019b. "An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features." *IEEE Transactions on Instrumentation and Measurement* 69 (4): 1493–1504. doi:10.1109/TIM.2019.2915404.

He, Z. Y., H. J. Wang, Y. Z. He, G. X. Zhang, J. Z. Wang, G. Y. Zou, and T. Chady. 2019. "Joint Scanning Laser Thermography Defect Detection Method for Carbon Fiber Reinforced Polymer." *IEEE Sensors Journal* 20 (1): 328–336. doi:10.1109/JSEN.2019.2941077.

Hinton, G. E., S. Osindero, and Y. W. The. 2006. "A Fast Learning Algorithm for Deep Belief Nets." *Neural Computation* 18 (7): 1527–1554. doi:10.1162/neco.2006.18.7.1527.

Hinton, G. E., O. Vinyals, and J. Dean. 2015. "Distilling the Knowledge in a Neural Network." arXiv preprint arXiv: 1503.02531, https://arxiv.org/abs/1503.02531.

Howard, A. G., M. Zhu, B. Chen, D. Kalenichenko, W. J. Wang, T. Weyand, M. Andreetto, and H. Adam. 2017. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications." arXiv preprint arXiv: 1704.04861, https://arxiv.org/abs/1704.04861.

Hu, C., and Y. Wang. 2019. "An Efficient CNN Model Based on Object-level Attention Mechanism for Casting Defects Detection on Radiography Images." *IEEE Transactions on Industrial Electronics* Early Access: 1. doi:10.1109/TIE.2019.2962437.

Hu, C. F., Y. X. Wang, K. Chen, Y. L. Qin, H. Shao, and J. K. Wang. 2019a. "A CNN Model Based on Spatial Attention Modules for Casting Type Classification on Pseudo-color Digital Radiography Images." In: Chinese automation congress (CAC 2019), Zhejiang, November 22–24. doi: 10.1109/CAC48633.2019.8996501.

Hu, J., L. Shen, S. Albanie, G. Sun, and E. H. Wu. 2019b. "Squeeze-and-Excitation Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* Early Access: 1. doi:10.1109/TPAMI.2019.2913372.

Hunag, G., S. C. Liu, L. V. D. Maaten, and K. Q. Weinberger. 2017. "CondenseNets: An Efficient DenseNet Using Learned Group Convolutions." arXiv preprint arXiv: 1711.09224, https://arxiv.org/abs/1711.09224.

Ibrahim, A., A. Dalbah, A. Abualsaud, U. Tariq, and A. El-Hag. 2020. "Application of Machine Learning to Evaluate Insulator Surface Erosion." *IEEE Transactions on Instrumentation and Measurement* 69 (2): 314–316. doi:10.1109/TIM.2019.2956300.

Ibrahim, M., M. Louie, C. Modarres, and J. Paisley. 2019. "Global Explanations of Neural Networks: Mapping the Landscape of Predictions." arXiv preprint arXiv: 1902.02384, https://arxiv.org/abs/1902.02384.

Imoto, K., T. Nakai, T. Ike, K. Haruki, and Y. Sato. 2018. "A CNN-Based Transfer Learning Method for Defect Classification in Semiconductor Manufacturing." In: International symposium on semiconductor manufacturing (ISSM 2018), Japan, December 10–11. doi: 10.1109/ISSM.2018.8651174.

Jian, C. X., J. Gao, and Y. H. Ao. 2017. "Automatic Surface Defect Detection for Mobile Phone Screen Glass Based on Machine Vision." *Applied Soft Computing* 52: 348–358. doi:10.1016/j.asoc.2016.10.030.

Jiao, W. H., Q. Y. Wang, Y. C. Cheng, and Y. M. Zhang. 2020. "End-to-end Prediction of Weld Penetration: A Deep Learning and Transfer Learning Based Method." *Journal of Manufacturing Processes*. doi:10.1016/j.jmapro.2020.01.044.

Kang, G. Q., S. B. Gao, L. Yu, and D. K. Zhang. 2018. "Deep Architecture for High-speed Railway Insulator Surface Defect Detection: Denoising Autoencoder with Multitask Learning."

IEEE Transactions on Instrumentation and Measurement 68 (8): 2679–2690. doi:10.1109/TIM.2018.2868490.

Kang, G. Q., S. B. Gao, L. Yu, D. K. Zhang, X. G. Wei, and D. Zhan. 2019. "Contact Wire Support Defect Detection Using Deep Bayesian Segmentation Neural Networks and Prior Geometric Knowledge." *IEEE Access* 7: 173366–173376. doi:10.1109/ACCESS.2019.2955753.

Kang, S. 2020. "Joint Modeling of Classification and Regression for Improving Faulty Wafer Detection in Semiconductor Manufacturing." *Journal of Intelligent Manufacturing* 31 (2): 319–326. doi:10.1007/s10845-018-1447-2.

Kendall, A., Y. Gal, and R. Cipolla. 2017. "Multi-task Learning Using Uncertainty to Weigh Losses for Scene Geometry and Semantics." arXiv preprint arXiv: 1705.07115v3, https://arxiv.org/abs/1705.07115v3.

Kendall, B., . A., R. Cipolla, and R. Cipolla. 2017. "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (12): 2481–2495. doi:10.1109/TPAMI.2016.2644615.

Kim, B., and F. Doshi-Velez. 2017. "Interpretable Machine Learning: The Fuss, the Concrete and the Questions." In: 34th international conference on machine learning (ICML 2017), Sydney, August 6–11. http://people.csail.mit.edu/beenkim/papers/BeenK_FinaleDV_ICML2017_tutorial.pdf.

Kim, M., M. Lee, M. An, and H. Lee. 2019. "Effective Automatic Sefect Classification Process Based on CNN with Stacking Ensemble Model for TFT-LCD Panel." *Journal of Intelligent Manufacturing* 31 (5): 1165–1174. doi:10.1007/s10845-019-01502-y.

Leary, J. O., K. Sawlani, and A. Mesbah. 2020. "Deep Learning for Classification of the Chemical Composition of Particle Defects on Semiconductor Wafers." *IEEE Transactions on Semiconductor Manufacturing* 33 (1): 72–85. doi:10.1109/TSM.2019.2963656.

Lecue, C. 2019. "On the Role of Knowledge Graphs in Explainable AI." *Semantic Web* 11 (1): 1–11. doi:10.3233/SW-190374.

Lecun, Y., Y. Bengio, and G. E. Hinton. 2015. "Deep Learning." *Nature* 521 (7553): 436–444. doi:10.1038/nature14539. Lecun, Y., L. Bottou, Y. Bengio, and P. Haffner. 1998. "Gradient-based Learning Applied to Document Recognition." Proceedings of the IEEE 86(11): 2278–2324. doi: 10.1109/5.726791.

Lee, D. H. 2013. "Pseudo-Label: The Simple and Efficient Semi-Supervised Learning Method for Deep Neural Networks." In: Workshop: challenges in representation learning (ICML 2013), Atlanta, Georgia, June 16–21.

Lei, T., Y. Huang, W. J. Shao, W. N. Liu, and Y. M. Rong. 2020a. "A Tactual Weld Seam Tracking Method in Super Narrow Gap of Thick Plates." *Robotics and Computer-Integrated Manufacturing* 62: 101864. doi:10.1016/j.rcim.2019.101864.

Lei, Y. X., X. F. Chen, M. C. Min, and Y. F. Xie. 2020b. "A Semi-supervised Laplacian Extreme Learning Machine and Feature Fusion with CNN for Industrial Superheat Identification." *Neurocomputing* 381: 186–195. doi:10.1016/j.neucom.2019.11.012.

Li, X., W. Li, Q. Yang, W. Yan, and A. Y. Zomaya. 2020. "An Unmanned Inspection System for Multiple Defects Detection in Photovoltaic Plants." *IEEE Journal of Photovoltaics* 10 (2): 568–576. doi:10.1109/JPHOTOV.2019.2955183.

Li, X., C. Ma, B. Wu, Z. He, and M. H. Yang. 2019. "Target-Aware Deep Tracking." In: IEEE conference on computer vision and pattern recognition (CVPR 2019), Long Beach, June 15–20. doi: 10.1109/CVPR.2019.00146.

Liang, J., C. Chen, J. Z. Liang, and Z. J. Hou. 2017. "Fabric Defect Inspection Based on Lattice Segmentation and Gabor Filtering." *Neurocomputing* 238: 84–102. doi:10.1016/j.neucom.2017.01.039.

Liang, Q. K., W. Zhu, W. Sun, Z. Yu, Y. N. Wang, and D. Zhang. 2019. "In-line Inspection Solution for Codes on Complex Backgrounds for the Plastic Container Industry." *Measurement* 148: 106965. doi:10.1016/j.measurement.2019.106965.

Lin, S., Z. He, and L. Sun. 2019. "Defect Enhancement Generative Adversarial Network for Enlarging Data Set of Microcrack Defect." *IEEE Access* 7: 148413–148423. doi:10.1109/ACCESS.2019.2946062.

Lin, S. F., T. Y. Wu, and Y. B. Hong. 2012. "X-ray Inspection Image Enhencement Using Shearlet Transform." In: IEEE conference on industrial electronics and applications (ICIEA 2012), Singapore, November 1763–1767. doi: 10.1109/ICIEA.2012.6361012.

Lipton, Z. C. 2018. "The Mythos of Model Interpretability." *Communications of the ACM* 61 (10): 36–43. doi:10.1145/3233231.

Liu, Q. W., X. Q. Xiang, Y. F. Wang, Z. W. Luo, and F. Fang. 2020a. "Aircraft Detection in Remote Sensing Image Based on Corner Clustering and Deep Learning." *Engineering Applications of Artificial Intelligence* 87: 103333. doi:10.1016/j.engappai.2019.103333.

Liu, T. Y., J. S. Bao, J. L. Wang, and Y. M. Zhang. 2020b. "A Coarse-Grained Regularization Method of Convolutional Kernel for Molten Pool Defect Identification." *Journal of Computing and Information Science in Engineering* 20 (2): 021105. doi:10.1115/1.4045294.

Liu, W., D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg. 2016. "SSD: Single Shot MultiBox Detector." In: European conference on computer vision (ECCV 2016), Amsterdam, October 8–16. doi: 10.1007/978-3-319-46448-0_2.

Loffe, S., and C. Szegedy. 2015. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift." arXiv preprint arXiv: 1502.03167, https://arxiv.org/abs/1502.03167.

Lu, C., L. Gao, X. Y. Li, W. Liao, and J. Yi. 2017. "An Effective Multi-objective Discrete Virus Optimization Algorithm for Flexible Job-shop Scheduling Problem with Controllable Processing Times." *Computers & Industrial Engineering* 104: 156–174. doi:10.1016/j.cie.2016.12.020.

Lu, C., L. Gao, Q. K. Pan, X. Y. Li, and J. Zheng. 2019. "A Multi-Objective Cellular Grey Wolf Optimizer for Hybrid Flowshop Scheduling Problem considering Noise

Pollution." *Applied Soft Computing* 75: 728–749. doi:10.1016/j.asoc.2018.11.043.

Lwahori, Y., Y. Takada, T. Shiina, M. K. Bhuyan, B. Kijsirikul, and B. Kijsirikul. 2018. "Defect Classification of Electronic Board Using Dense SIFT and CNN." *Procedia Computer Science* 126: 1673–1682. doi:10.1016/j.procs.2018.08.110.

Ma, C., J. B. Huang, X. Yang, and M. H. Yang. 2015. "Hierarchical Convolutional Features for Visual Tracking." In: IEEE international conference on computer vision (ICCV 2015), Santiago, December 7–13. doi: 10.1109/ICCV.2015.352.

Malarvel, M., G. Sethumadhavan, P. C. R. Bhagi, S. Kar, T. Saravanan, and A. Krishnan. 2017. "Anisotropic Diffusion Based Denoising on X-radiography Images to Detect Weld Defects." *Digital Signal Processing* 68: 112–126. doi:10.1016/j.dsp.2017.05.014.

Mei, S., Y. D. Wang, and G. J. Wen. 2018. "Automatic Fabric Defect Detection with a Multi-Scale Convolutional Denoising Autoencoder Network Model." *Sensors* 18 (4): 1064. doi:10.3390/s18041064.

Meijer, D., L. Scholten, F. Clemens, and A. Knobbe. 2019. "A Defect Classification Methodology for Sewer Image Sets with Convolutional Neural Networks." *Automation in Construction* 104: 281–298. doi:10.1016/j.autcon.2019.04.013.

Michalski, P., S. Zator, B. Ruszczak, and M. Tomaszewski. 2019. "Detection of Power Line Insulators on Digital Images with the Use of Laser Spots." *IET Image Processing* 13 (12): 2358. doi:10.1049/iet-ipr.2018.6284.

Ming, W. Y., F. S. Shen, H. M. Zhang, X. K. Li, J. Ma, J. G. Du, and Y. Lu. 2019. "Defect Detection of LGP Based on Combined Classifier with Dynamic Weights." *Measurement* 143: 211–225. doi:10.1016/j.measurement.2019.04.087.

Nam, H., and B. Han. 2016. "Learning Multi-domain Convolutional Neural Networks for Visual Tracking." In: IEEE conference on computer vision and pattern recognition (CVPR 2016), Las Vegas, June 27–30. doi: 10.1109/CVPR.2016.465.

Niu, S., B. Li, X. Wang, and H. Lin. 2020. "Defect Image Sample Generation with GAN for Improving Defect Recognition." *IEEE Transactions on Automation Science and Engineering* 99: 1–12. doi:10.1109/TASE.2020.2967415.

Pan, S. J., and Y. Qiang. 2010. "A Survey on Transfer Learning." *IEEE Transactions on Knowledge and Data Engineering* 22 (10): 1345–1359. doi:10.1109/TKDE.2009.191.

Park, S., and N. Kwak. 2016. "Analysis on the Dropout Effect in Convolutional Neural Networks." In: Asian conference on computer vision. springer (ACCV 2016), Taipei, November 20–24. doi: 10.1007/978-3-319-54184-6_12.

Racki, D., D. Tomazevic, and D. Skocaj. 2018. "A Compact Convolutional Neural Network for Textured Surface Anomaly Detection." In: IEEE winter conference on applications of computer vision (WACV 2018), Lake Tahoe, March 12–15. doi: 10.1109/WACV.2018.00150.

Redmon, J., S. Divvala, R. Girshick, and A. Farhadi. 2016. "You Only Look Once: Unified, Real-Time Object Detection." In: IEEE conference on computer vision and pattern recognition (CVPR 2016), Las Vegas, June 27–30. doi: 10.1109/CVPR.2016.91.

Ren, S. Q., K. He, R. Girshick, and J. Sun. 2016. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6): 1137–1149. doi:10.1109/TPAMI.2016.2577031.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2016. "Why Should I Trust You?": Explaining the Predictions of Any Classifier." arXiv preprint arXiv: 1602.04938, https://arxiv.org/abs/1602.04938.

Ribeiro, M. T., S. Singh, and C. Guestrin. 2018. "Anchors: High-precision Model-agnostic Explanations." In: 32nd AAAI conference on artificial intelligence (AAAI 2018), Louisiana, February 2–7. https://homes.cs.washington.edu/~marcotcr/aaai18.pdf

Ronneberger, O., P. Fischer, and T. Brox. 2015. "U-Net: Convolutional Networks for Biomedical Image Segmentation." In: 18th international conference on medical image computing and computer assisted intervention (MICCAI 2015), Munich, October 5–9. doi: 10.1007/978-3-319-24574-4_28.

Sabour, S., N. Frosst, and G. E. Hinton. 2017. "Dynamic Routing Between Capsules." arXiv preprint arXiv: 1710.09829, https://arxiv.org/abs/1710.09829.

Scime, L., and J. Beuth. 2018. "A Multi-scale Convolutional Neural Network for Autonomous Anomaly Detection and Classification in A Laser Powder Bed Fusion Additive Manufacturing Process." *Additive Manufacturing* 24: 273–286. doi:10.1016/j.addma.2018.09.034.

Shehab, M., L. Abualigah, M. I. Jarrah, O. A. Alomari, and M. S. Daoud. 2020. "Artificial Intelligence in Software Engineering and Inverse: Review." *International Journal of Computer Integrated Manufacturing* 33(10-11 (10–11): 1129–1144. doi:10.1080/0951192X.2020.1780320.

Shelhamer, E., J. Long, and T. Darrell. 2017. "Fully Convolutional Networks for Semantic Segmentation." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (4): 640–651. doi:10.1109/TPAMI.2016.2572683.

Shen, X. H., Z. H. Li, M. Li, X. L. Xu, and X. W. Zhang. 2020. "Aluminum Surface Defect Detection Based on Multi-task Deep Learning." *Laser & Optoelectronics Progress* 57 (10): 101501. doi:10.3788/lop57.101501.

Shilon, I., M. Kraus, M. Büchele, K. Egberts, T. Fischer, T. L. Holch, T. Lohse, U. Schwanke, C. Steppa, and S. Funk. 2019. "Application of Deep Learning Methods to Analysis of Imaging Atmospheric Cherenkov Telescopes Data." *Astroparticle Physics* 105: 44–53. doi:10.1016/j.astropartphys.2018.10.003.

Simonyan, K., A. Vedaldi, and A. Zisserman. 2013. "Deep inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps." arXiv preprint arXiv:1312.6034, https://arxiv.org/abs/1312.6034.

Simonyan, K., and A. Zisserman. 2015. "Very Deep Convolutional Networks for Large-Scale Image Recognition." In: 3rd international conference on learning representations (ICLR 2015), San Diego, May 7–9. arXiv preprint arXiv: 1409.1556, https://arxiv.org/abs/1409.1556.

Song, G. R., K. C. Song, and Y. H. Yan. 2020. "Saliency Detection for Strip Steel Surface Defects Using Multiple Constraints and Improved Texture Features." *Optics and Lasers in Engineering* 128: 106000. doi:10.1016/j.optlaseng.2019.106000.

Song, K. C., and Y. H. Yan. 2013. "A Noise Robust Method Based on Completed Local Binary Patterns for Hot-rolled Steel Strip Surface Defects." *Applied Surface Science* 285: 858–864. doi:10.1016/j.apsusc.2013.09.002.

Springenberg, J. T., A. Dosovitskiy, T. Brox, and M. Riedmiller. 2014. "Striving for Simplicity: The All Convolutional Net." arXiv preprint arXiv:1412.6806, https://arxiv.org/abs/1412.6806.

Sung, A. H. 1998. "Ranking Importance of Input Parameters of Neural Networks." *Expert Systems with Applications* 15 (3–4): 405–411. doi:10.1016/s0957-4174(98)00041-4.

Szegedy, C., W. Liu, Y. Q. Jia, P. Sermanet, S. Reed, and D. Anguelov. 2015. "Going Deeper with Convolutions." In: IEEE conference on computer vision and pattern recognition (CVPR 2015), Boston, June 7–12. doi: 10.1109/CVPR.2015.7298594.

Tabernik, D., S. Šela, J. Skvarč, and D. Skočaj. 2020. "Segmentation-based Deep-learning Approach for Surface-defect Detection." *Journal of Intelligent Manufacturing* 31 (3): 759–776. doi:10.1007/s10845-019-01476-x.

Tang, W. Q., Q. Yang, K. X. Xiong, and W. J. Yan. 2020. "Deep Learning Based Automatic Defect Identification of Photovoltaic Module Using Electroluminescence Images." *Solar Energy* 201: 453–460. doi:10.1016/j.solener.2020.03.049.

Tao, X., D. P. Zhang, W. Z. Ma, X. L. Liu, and D. Xu. 2018. "Automatic Metallic Surface Defect Detection and Recognition with Convolutional Neural Networks." *Applied Science* 8 (9): 1575. doi:10.3390/app8091575.

Tian, S. S., Y. X. Li, Y. L. Kang, and J. N. Xia. 2021. "Multi-robot Path Planning in Wireless Sensor Networks Based on Jump Mechanism PSO and Safety Gap Obstacle Avoidance." *Future Generation Computer Systems* 118: 37–47. doi:10.1016/j.future.2020.12.012.

Tompson, J., R. Goroshin, A. Jain, Y. Lecun, and C. Bregler. 2015. "Efficient Object Localization Using Convolutional Networks." In: IEEE conference on computer vision and pattern recognition (CVPR 2015), Boston, June 7–12. doi: 10.1109/CVPR.2015.7298664.

Valavanis, I., and D. Kosmopoulos. 2010. "Multiclass Defect Detection and Classification in Weld Radiographic Images Using Geometric and Texture Features." *Expert Systems with Applications* 37 (12): 7606–7614. doi:10.1016/j.eswa.2010.04.082.

Wang, H. J., S. J. Hsieh, X. F. Zhou, P. Bo, and B. Singh. 2015. "Using Active Thermography to Inspect Pin-hole Defects in Anti-reflective Coating with K-mean Clustering." *NDT & E International* 76: 66–72. doi:10.1016/j.ndteint.2015.08.006.

Wang, J. J., P. L. Fu, and R. X. Gao. 2019. "Machine Vision Intelligence for Product Defect Inspection Based on Deep Learning and Hough Transform." *Journal of Manufacturing Systems* 51: 52–60. doi:10.1016/j.jmsy.2019.03.002.

Wang, J. J., Y. L. Ma, L. B. Zhang, R. X. Gao, and D. Z. Wu. 2018. "Deep Learning for Smart Manufacturing: Methods and Applications." *Journal of Manufacturing Systems* 48 (Part C): 144–156. doi:10.1016/j.jmsy.2018.01.003.

Wang, J. L., C. Q. Xu, L. Dai, J. Zhang, and R. Y. Zhong. 2020a. "An Unequal Learning Approach for 3D Point Cloud Segmentation." *IEEE Transactions on Industrial Informatics* Early Access: 1. doi:10.1109/TII.2020.3044106.

Wang, J. L., C. Q. Xu, Z. L. Yang, J. Zhang, and X. O. Li. 2020b. "Deformable Convolutional Networks for Efficient Mixed-Type Wafer Defect Pattern Recognition." *IEEE Transactions on Semiconductor Manufacturing* 33 (4): 587–596. doi:10.1109/TSM.2020.3020985.

Wang, J. L., Z. L. Yang, J. Zhang, Q. H. Zhang, and W. K. Chien. 2019a. "AdaBalGAN: An Improved Generative Adversarial Network with Imbalanced Learning for Wafer Defective Pattern Recognition." *IEEE Transactions on Semiconductor Manufacturing* 32 (3): 310–319. doi:10.1109/TSM.2019.2925361.

Wang, R., and N. Chen. 2019. "Wafer Map Defect Pattern Recognition Using Rotation-Invariant Features." *IEEE Transactions on Semiconductor Manufacturing* 32 (4): 596–604. doi:10.1109/TSM.2019.2944181.

Wang, T., J. Q. Chen, X. D. Gao,and, and Y. X. Qin. 2017. "Real-time Monitoring for Disk Laser Welding Based on Feature Selection and SVM." *Applied Sciences* 7 (9): 884. doi:10.3390/app7090884.

Wang, Y. M., H. M. Du, X. Zhang, and Y. D. Xu. 2019b. "Application of Visual Attention Network in Workpiece Surface Defect Detection." *Journal of Computer-Aided Design & Computer Graphics* 31 (9): 1528–1534. doi:10.3724/SP.J.1089.2019.17506.

Wei, X. K., Z. M. Yang, X. Y. Liu, D. H. Wei, L. M. Jia, and Y. J. Li. 2019. ""Railway Track Fastener Defect Detection Based on Image Processing and Deep Learning Techniques: A Comparative Study." *Engineering Applications of Artificial Intelligence* 80: 66–81. doi:10.1016/j.engappai.2019.01.008.

Woo, S., J. Park, J. Y. Lee, and I. S. Kweon. 2018. "CBAM: Convolutional Block Attention Module." arXiv preprint arXiv: 1807.06521, https://arxiv.org/abs/1807.06521.

Wu, H., and X. Gu. 2015. "Towards Dropout Training for Convolutional Neural Networks." *Neural Networks* 71 (C): 1–10. doi:10.1016/j.neunet.2015.07.007.

Wu, J. X., N. A. Cai, F. Y. Li, H. W. Jiang, and H. Wang. 2020a. "Automatic Detonator Code Recognition via Deep Neural Network." *Expert Systems with Applications* 145: 113121. doi:10.1016/j.eswa.2019.113121.

Wu, M., J. R. Jang, and J. Chen. 2015. "Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-scale Data Sets." *IEEE Transactions on Semiconductor Manufacturing* 28 (1): 1–12. doi:10.1109/TSM.2014.2364237.

Wu, Y. P., Y. Qin, Z. P. Wang, X. P. Ma, and Z. W. Cao. 2020b. "Densely Pyramidal Residual Network for UAV-based Railway Images Dehazing." *Neurocomputing* 371: 124–136. doi:10.1016/j.neucom.2019.06.076.

Xia, B. X., J. B. Cao, and C. Wang. 2019. "SSIM-NET: Real-time PCB Defect Detection Based on SSIM and MobileNet-V3." In: 2nd world conference on mechanical engineering and intelligent manufacturing (WCMEIM 2019), Shanghai, November 22–24. doi: 10.1109/WCMEIM48965.2019.00159.

Xiang, Z., K. F. Chen, M. Qian, and X. D. Hu. 2019. "Yarn-dyed Woven Fabric Density Measurement Method and System Based on Multi-Directional Illumination Image Fusion Enhancement Technology." *Journal of the Textile Institute*. doi:10.1080/00405000.2019.1706222.

Xiao, H. F., D. Chen, J. W. Xu, and S. F. Guo. 2020. "Defects Identification Using the Improved Ultrasonic Measurement Model and Support Vector Machines." *NDT & E International* 111: 102223. doi:10.1016/j.ndteint.2020.102223.

Xie, L. F., X. Xiang, H. N. Xu, L. Wang, L. J. Lin, and G. F. Yin. 2020. "FFCNN: A Deep Neural Network for Surface Defect Detection of Magnetic Tile." *IEEE Transactions on Industrial Electronics* Early Access: 1. doi:10.1109/TIE.2020.2982115.

Xie, Q., D. W. Li, J. X. Xu, Z. H. Yu, and J. Wang. 2019. "Automatic Detection and Classification of Sewer Defects via Hierarchical Deep Learning." *IEEE Transactions on Automation Science and Engineering* 16 (4): 1836–1847. doi:10.1109/TASE.2019.2900170.

Xu, L., S. Lv, Y. Deng, and X. X. Li. 2020. "A Weakly Supervised Surface Defect Detection Based on Convolutional Neural Network." *IEEE Access* 8: 42285–42296. doi:10.1109/access.2020.2977821.

Xu, T. Y., Z. H. Feng, X. J. Wu, and J. Kittler. 2019. "Learning Adaptive Discriminative Correlation Filters via Temporal Consistency Preserving Spatial Feature Selection for Robust Visual Tracking." *IEEE Transactions on Image Processing* 28 (11): 5596–5609. doi:10.1109/TIP.2019.2919201.

Yang, L. J., Z. J. Wang, S. W. Gao, M. Shi, and B. L. Liu. 2019. "Magnetic Flux Leakage Image Classification Method for Pipeline Weld Based on Optimized Convolution Kernel." *Neurocomputing* 365: 229–238. doi:10.1016/j.neucom.2019.07.083.

Yang, X., M. R. Han, H. L. Tang, Q. Li, and X. Luo. 2020. "Detecting Defects with Support Vector Machine in Logistics Packaging Boxes for Edge Computing." *IEEE Access* 8: 64002–64010. doi:10.1109/ACCESS.2020.2984539.

Yu, X. L., X. Lin, Y. Q. Dai, and K. Zhu. 2017. "Image Edge Detection Based Tool Condition Monitoring with Morphological Component Analysis." *ISA Transactions* 69: 315–322. doi:10.1016/j.isatra.2017.03.024.

Yu, Y. J., H. Cao, X. Y. Yan, T. Wang, and S. S. Ge. 2019. "Defect Identification of Wind Turbine Blades Based on Defect Semantic Features with Transfer Feature Extractor." *Neurocomputing* 376: 1–9. doi:10.1016/j.neucom.2019.09.071.

Zeng, Z., S. G. Dai, and P. A. Mu. 2007. "Wafer Defects Detecting and Classifying System Based on Machine Vision." In: International conference on electronic measurement and instruments, Xi'an, October 520–523. doi: 10.1109/ICEMI.2007.4351196.

Zhang, D. F., K. C. Song, J. Xu, Y. He, and Y. H. Yan. 2019a. ""Unified Detection Method of Aluminium Profile Surface Defects: Common and Rare Defect Categories." *Optics and Lasers in Engineering* 126: 105936. doi:10.1016/j.optlaseng.2019.105936.

Zhang, H., J. Ma, J. Jing, and P. Li. 2019b. "Fabric Defect Detection Using L0 Gradient Minimization and Fuzzy C-Means." *Applied Science* 9 (17): 3506. doi:10.3390/app9173506.

Zhang, J. J., P. Wang, R. Q. Yan, and R. X. Gao. 2018a. "Long Short-term Memory for Machine Remaining Life Prediction." *Journal of Manufacturing Systems* 48 (Part C): 78–86. doi:10.1016/j.jmsy.2018.05.011.

Zhang, Q. S., W. G. Wang, and S. C. Zhu. 2017. "Examining CNN Representations with respect to Dataset Bias." arXiv preprint arXiv: 1710.10577, https://arxiv.org/abs/1710.10577.

Zhang, W. J., G. Yang, Y. Lin, C. Ji, and M. M. Gupta. 2018b. "On Definition of Deep Learning." In: IEEE 2018 world automation congress (WAC 2018), Stevenson, June 3–6. doi: 10.23919/WAC.2018.8430387.

Zhang, X. Y., X. Y. Zhou, M. X. Lin, and J. Sun. 2017. "ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices." arXiv preprint arXiv: 1707.01083, https://arxiv.org/abs/1707.01083.

Zhang, Y. X., D. Y. You, X. D. Gao, N. F. Zhang, and P. P. Gao. 2019c. "Welding Defects Detection Based on Deep Learning with Multiple Optical Sensors during Disk Laser Welding of Thick Plates." *Journal of Manufacturing Systems* 51: 87–94. doi:10.1016/j.jmsy.2019.02.004.

Zhang, Z. F., W. J. Ren, Z. Yang, and G. R. Wen. 2020a. "Real-time Seam Defect Identification for Al Alloys in Robotic Arc Welding Using Optical Spectroscopy and Integrating Learning." *Measurement* 156: 107546. doi:10.1016/j.measurement.2020.107546.

Zhang, Z. F., G. R. Wen, and S. B. Chen. 2019. "Weld Image Deep Learning-based On-line Defects Detection Using Convolutional Neural Network for Al Alloy in Robotic Arc Welding." *Journal of Manufacturing Processes* 45: 208–216. doi:10.1016/j.jmapro.2019.06.023.

Zhang, Z. H., B. Li, W. F. Zhang, R. D. Lu, S. Wada, and Y. Zhang. 2020b. "Real-time Penetration State Monitoring Using Convolutional Neural Network for Laser Welding of Tailor Rolled Blanks." *Journal of Manufacturing Systems* 54: 348–360. doi:10.1016/j.jmsy.2020.01.006.

Zhao, H. S., J. P. Shi, X. J. Qi, X. G. Wang, and J. Y. Jia. 2017. "Pyramid Scene Parsing Network." In: IEEE conference on computer vision and pattern recognition (CVPR 2017), Honolulu, July 21–26. doi: 10.1109/CVPR.2017.660.

Zhao, J. L., Y. S. Peng, and Y. H. Yan. 2018. "Steel Surface Defect Classification Based on Discriminant Manifold Regularized Local Descriptor." *IEEE Access* 6: 71719–71731. doi:10.1109/ACCESS.2018.2881962.

Zhao, Y. D., K. R. Hao, H. B. He, X. S. Tang, and B. Wei. 2020. "A Visual Long-hort-term Memory Based Integrated CNN Model for Fabric Defect Image Classification." *Neurocomputing* 380: 259–270. doi:10.1016/j.neucom.2019.10.067.

Zhou, B. L., A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. 2016. "1Learning Deep Features for Discriminative Localization." In: IEEE conference on computer vision and pattern recognition (CVPR 2016), Las Vegas, June 27–30. doi: 10.1109/CVPR.2016.319.

Zhou, P., G. B. Zhou, Y. M. Li, Z. Z. He, and Y. W. Liu. 2020. "A Hybrid Data-Driven Method for Wire Rope Surface Defect Detection." *IEEE Sensors Journal* 99: 1. doi:10.1109/JSEN.2020.2970070.

Zhou, Z. H. 2016. *Machine Learning*. Beijing: Tsinghua University Press.

Zhuang, J. T., and J. L. Yang. 2018. "ShelfNet for Real-time Semantic Segmentation." arXiv preprint arXiv: 1811.11254v2, https://arxiv.org/abs/1811.11254v2.