



# A novel deep learning motivated data augmentation system based on defect segmentation requirements

Shuanlong Niu<sup>1</sup> · Yaru Peng<sup>1</sup> · Bin Li<sup>1</sup> · Yuanhong Qiu<sup>1</sup> · Tongzhi Niu<sup>1</sup> · Weifeng Li<sup>1</sup>

Received: 3 May 2022 / Accepted: 7 December 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

## Abstract

Deep learning methods, especially convolutional neural networks (CNNs), are widely used for industrial surface defect segmentation due to their excellent performance on visual inspection tasks. However, the problems of overfitting and low generalizability affect the performance of CNN-based surface defect segmentation models. Therefore, data augmentation is necessary to reduce overfitting and improve generalization. However, existing data augmentation methods are random and independent of the downstream defect detection tasks. Thus, we propose a simple plug-and-play data augmentation method based on the requirements of the CNN defect segmentation task. We first pretrain a defect segmentation model on a training set and obtain confidence maps. Then, we occlude high-confidence regions based on the data augmentation module to balance the attention paid by the model to high- and low-confidence regions. Finally, to prevent overfitting, we periodically update the confidence map. We conduct sufficient experiments on the Kolektor surface defect (KSD) metal surface dataset and an optoelectronic chip dataset. The proposed method can make full use of the information contained in the existing datasets to improve the accuracy of the segmentation model (intersection-over-union (IOU) 6.62% improvement over baseline on the KSD dataset and 3.20% on the optoelectronic chip dataset). Moreover, the proposed method is highly applicable to various mainstream defect segmentation methods.

**Keywords** Defect detection · Image segmentation · Data augmentation · Deep learning

## Introduction

Industrial visual inspection is very important for the intellectualization and automation of manufacturing, and it is a key technology for process control and quality inspection in industrial production lines. Recently, due to remarkable deep learning results, especially those of convolutional neural networks (CNNs), in the computer vision field, applying CNNs in the vision inspection field for various industrial products, such as metals (Chuanfei & Wang, 2020), textiles (Dong et al., 2020), solar cells (Su et al., 2021) and wafer (Kim et al., 2021; Cheng et al., 2021), has become a research hotspot.

However, most CNN models are notoriously data-hungry because of their massive numbers of parameters (Chen et al., 2020). It is difficult for industrial sites to collect defect images with high information density to build sufficient defect image datasets to meet the needs of deep learning. Therefore, the problems of overfitting and low generalizability (Krizhevsky et al., 2012) may arise in CNN-based surface defect detection.

Data augmentation has become a feasible technology for mitigating overfitting and improving generalization. On the one hand, many researchers have generated new images based on generative adversarial networks (GANs) (Goodfellow et al., 2014), variational autoencoders (VAEs) (Kingma & Welling, 2013), and their variants (Isola et al., 2017; Choi et al., 2018) to increase the diversity of datasets and address generalization problems. The idea of generative data augmentation methods has also been successfully applied to pedestrian detection (Zheng et al., 2017), license plate recognition (Wang et al., 2017), road recognition (Fan et al., 2021), etc. However, the training of generative models requires a

✉ Bin Li  
libin999\_hust@163.com  
Shuanlong Niu  
d201780187@hust.edu.cn

<sup>1</sup> Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Luoyu Road 1037, Wuhan 430074, Hubei, China

certain amount of data, which is difficult to obtain in many industrial scenarios. On the other hand, many researchers have proposed information dropping-based data augmentation methods, such as Cutout (DeVries & Taylor, 2017), which randomly deletes a part of the input image to generate a new image for training, thereby increasing the robustness of the model. However, the existing data augmentation methods still have the following problems.

1. The above methods are random. For example, the image generated by a GAN (Goodfellow et al., 2014) is based on random hidden variables. The region deleted by Cutout (DeVries & Taylor, 2017) is also random. Randomness results in invalid augmented images, reducing augmentation efficiency.
2. The augmentation task itself is separated from the downstream detection task. As shown in Fig. 1, the pictured defect segmentation model did not have any effect on data augmentation.

However, the purpose of data augmentation is to improve the performance of downstream CNN tasks. Therefore, the efficiency and effectiveness of data augmentation will be significantly improved based on the requirements of the given defect segmentation task.

This paper focuses on data augmentation for defect segmentation. Defect segmentation can predict the category for each pixel of the industrial image, so the accurate defect area can be obtained. Therefore, certain key features can be obtained by segmentation results, including position, shape, area, and contour, which are used to accurately judge the size and severity of defects. The mainstream CNN segmentation model includes a feature extractor and a pixel classifier. The confidence map shown in Fig. 2 can be obtained by the feature extractor, and this map represents the confidence values of the CNN for different defect regions. The CNN model has high confidence for the defect region in the yellow box, but it does not have sufficient confidence to detect the defect region in the red box in Fig. 2. Therefore, the proposed data augmentation approach shown in Fig. 3 improves the attention and ability of the CNN to detect low-confidence areas. First, we pretrain the CNN model without data augmentation to reach a steady state. Then, the confidence map is obtained through the feature extractor. The defect regions with high confidence can be occluded, while those with low confidence can be maintained, forcing the CNN to focus on the defect regions that are difficult to detect. Finally, the confidence map can be updated to prevent overfitting after a period of training. Compared with other industrial data augmentation approaches, the proposed method does not need to train additional image generation models. In addition, it is simple to operate as a separate module that can be added to various existing defect segmentation models. Moreover,

the proposed method fully utilizes the information contained in the existing dataset to meet the requirements of the CNN model. With the proposed data augmentation technique, the CNN model can significantly improve its defect segmentation accuracy and robustness, especially for defect regions with low confidence.

The contributions can be summarized as follows:

1. We propose a simple, plug-and-play, and effective data augmentation method. Based on the confidence map obtained through pretraining, our approach can delete high-confidence defect regions to balance the attention paid by the defect segmentation model to high- and low-confidence regions.
2. We propose a novel defect augmentation approach that performs defect augmentation according to the requirements of the downstream defect segmentation model.
3. The proposed data augmentation method proves its effectiveness in terms of achieving improved defect segmentation accuracy and robustness and is highly generalizable to different existing segmentation frameworks.

The remainder of the article is organized as follows. The related work on data augmentation methods and their applications in industrial vision is discussed in “[The related work](#)” section. “[Methodology](#)” section describes the proposed method in detail. “[Experiments](#)” section presents the experimental results in comparison with those of different state-of-the-art methods. Finally, the conclusion and future work are summarized in “[Conclusion](#)” section.

## The related work

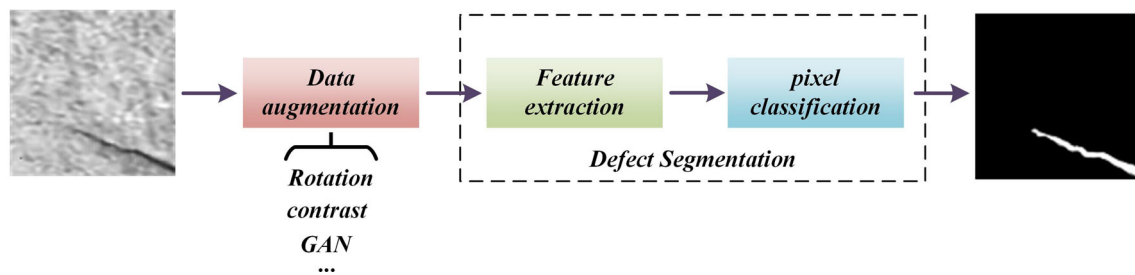
### Data augmentation

Data augmentation is an important way to solve data shortages and CNN model overfitting problems, and augmentation approaches can be divided into image conversion, image fusion, and image generation techniques.

Image conversion-based data augmentation methods include spatial transformations and color transformations.

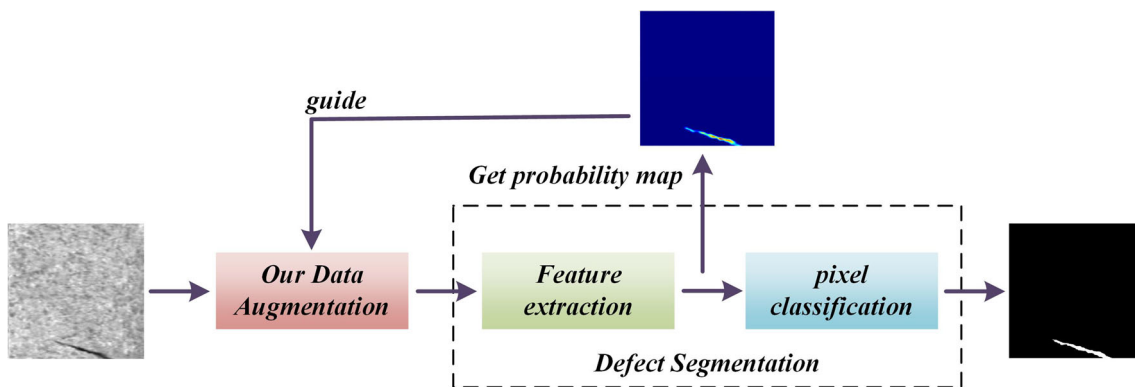
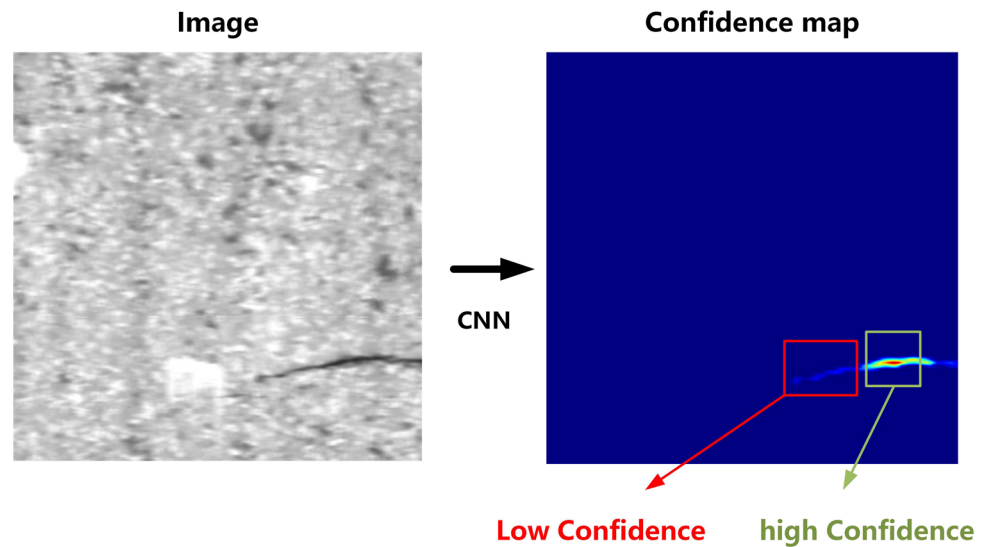
1. Spatial transformations: flipping, translation, rotation, and so on.
2. Color transformations: noise injection, contrast, brightness transformation, and so on.

Image conversions have become the basic operations in visual tasks involving various CNNs. For example, Müller and Hutter (2021) proposed TrivialAugment, which can sample an augmentation and a strength for each image and apply it. Recently, information dropping has become a new



**Fig. 1** The pipeline of traditional data augmentation methods

**Fig. 2** The confidence map for defect images obtained by a CNN



**Fig. 3** The proposed data augmentation method guided by a CNN

and popular research topic for image conversion-based data augmentation. DeVries and Taylor (2017) proposed Cutout, which randomly uses a square mask as a regularization method to occlude a part of the input image. HAS (Singh & Lee, 2017) divides an image into  $S \times S$  blocks and randomly hides some image blocks. Chen et al. (2020) proposed a structured occlusion-based augmentation method named GridMask that can delete regions with a set of spatially and uniformly distributed squares.

Image fusion combines the information derived from two or more images to form a new image. Zhang et al. (2017) proposed Mixup, which performs weighted pixel interpolation with two images and their labels to obtain new samples. Similar to Cutout, Cutmix (Yun et al., 2019) first occludes parts of one image and then fills the contents of another image with the corresponding position to achieve the fusion of the two images. Hendrycks et al. (2019) proposed Augmix, which first uses geometric and color transformation to augment the input image and then fuses the augmented images. TransMix

(Chen et al., 2021) fuses images and labels based on attention guidance for vision transformation tasks. Ghiasi et al. (2021) simply copy-pastes objects onto random images to augment instance segmentation datasets.

Image generation is mainly based on the variants of VAEs or GANs and generates a large number of images to increase the volume and diversity of datasets. It is widely used in different scenarios, such as pedestrian reidentification, road recognition, medical image detection and face recognition. Zheng et al. (2019) transformed pedestrian attributes and styles through a GAN to perform data augmentation and thus improved the accuracy of pedestrian reidentification. Lin et al. (2020) improved the robustness of a road detection model by mapping road images obtained during the daytime and good weather to night, rainy and foggy weather conditions. Uricar et al. (2021) proposed a dirty image GAN-based generation method to improve the diversity of the wide-angle fisheye camera dataset and significantly improve the autopilot accuracy of harsh environmental settings. Choi et al. (2019) presented a GAN-based data augmentation method to facilitate domain alignment between the images of game scenes and real scenes, reducing the difficulty and cost of data collection. Zhang et al. (2020) transformed license plate images produced by a template into a real scene through CycleGAN (Zhu et al., 2017) to improve the accuracy of license plate detection in unconstrained environments, such as reflection, blurring, and tilting. Bissoto et al. (2021) synthesized samples indistinguishable from real images for skin lesion medical applications based on GANs, obtained favorable results from out-of-distribution test sets and achieved anonymization. Naghizadeh et al. (2021) proposed a data augmentation method for biomedical and natural images, which can find and select the best policies for an autoencoder through a greedy search to generate cell images and masks. Luo et al. (2021) presented a face augmentation generative adversarial network (FA-GAN) that designs a hierarchical disentanglement module to decouple the attributes of identity representation and can be applied to long-tail imbalanced data distributions.

## Industrial data augmentation

Due to the characteristics of high yield rates and small numbers of abnormal samples in industrial sites, it is difficult to construct an industrial defect dataset that meets the requirements of deep learning for building CNN models with good generalization and robustness. Therefore, many scholars have attempted to design image augmentation methods for industrial defect detection (Hsu & Liu, 2021; Schlosser et al., 2022).

Many researchers have used geometric and color transformations of existing images to expand defect datasets. Sassi et al. (2019) performed random rotations and flips to

limit the decrease in performance for injector defect detection. Pan et al. (2021) used traditional image transformation methods, such as rotation, cropping, flipping, and brightness and contrast adjustment, to augment CT defect images of wood. Recently, with the rapid development of image generation technology, such as GANs, many researchers have tried to apply these approaches to industrial defect image generation (Jain et al., 2020; Xuan et al., 2018). Niu et al. (2020) proposed the surface defect-generation adversarial network (SD-GAN) based on a cycle-consistent loss and a dual adversarial loss to generate new defects with balanced image quality and diversity based on a large number of defect-free images. After generating defect images via Wasserstein GANs (WGANs), Le et al. (2020) constructed a quality management model with fake and real images by transfer learning. Yang et al. (2021) proposed a new data augmentation algorithm, named Mask2Defect, via prior knowledge infusion, which was able to generate a large volume of defects with different shapes, severities, scales, rotation angles, spatial locations, and part numbers. Yun et al. (2020) proposed a conditional CVAE (CCVAE) to generate images of each defect type in a single CVAE model for training a classification model. Liu et al. (2019) proposed a multistage GAN to generate different defect patches given different texture backgrounds and then fused the generated defects at specific locations. Liu et al. (2021) proposed a novel stacked multimanifold autoencoder (S-MMAE) by introducing a new multimanifold regularization to offer a more comprehensive representation of original data, which can significantly increase prediction accuracy. Based on a GAN and the Gaussian mixture model, Ren et al. (2022) proposed a data augmentation method that can effectively increase the number of training samples and reduce distribution differences between training and test sets and thus obtain a significant improvement in sanitary-ceramic defect classification. Li et al. (2022) proposed a new augmentation model named extremely imbalanced data augmentation generative adversarial network (EID-GAN) with a new penalty function that guides the generator to learn the features of outliers, thus addressing the extremely imbalanced data augmentation problem. Selecting the conditional GAN and geometrical transformation techniques, Liu et al. (2021) proposed a data augmentation method to improve the diversity and realism of synthetic defects in laser line scan sensor images. Tang et al. (2022) proposed a data generation methodology based on homography transformation and image fusion that can generate tobacco pack images with foreign objects. The above defect augmentation methods can significantly improve the performance of the downstream task defect detection model. However, they require a certain amount of data to build a generative model and are independent of downstream task defect detection.

## Methodology

### The pipeline of the proposed method

In this section, we propose a new data augmentation method based on the requirements of downstream CNN-based defect segmentation tasks shown in Fig. 4, which contains the following modules.

1. Baseline defect segmentation module: This module includes two submodules: feature extractor and pixel classifier. The feature extractor can be used to obtain confidence maps for defect images in training sets.
2. Data augmentation module based on a CNN: The information-dropping data augmentation method is performed according to the obtained confidence maps.
3. Loss function module: This module can be divided into normal loss functions before data augmentation and weight-guided loss functions during data augmentation.

The proposed data augmentation strategy is a simple and plug-and-play method that can be directly applied to existing CNN-based defect segmentation algorithms during training. Therefore, the training process of the proposed method can be divided into a normal mode and an augmentation mode. The normal mode used in the early stage trains the defect segmentation model without data augmentation. Until the model reaches a steady state with a stable confidence map, we use the augmentation mode to further improve the performance of the CNN model.

### The baselines for CNN defect segmentation

Existing defect segmentation models generally use popular image segmentation algorithms, such as UNet (Ronneberger et al., 2015), DeepLab (Chen et al., 2018), and fully convolutional networks (FCNs) (Long et al., 2015), which are widely utilized in various defect segmentation scenarios. The defect segmentation model  $f_s$  can generally be divided into feature extractor and pixel classifier modules.

Feature extractor module  $f_{fe}$ : consists of different network modules, such as the inception module (Szegedy et al., 2016), resblock module (He et al., 2016), and dense module (Huang et al., 2017), which are composed of different base layers, such as convolution, pooling, and normalization layers. It mainly extracts the features of input image  $X$  in a layer-by-layer manner to obtain confidence maps  $P$ .

$$P = f_{fe}(X) \quad (1)$$

Pixel classifier module  $f_{pc}$ : consists of Sigmoid or Softmax layers. It can determine image segmentation results  $R$  based

on the obtained confidence map  $P$ .

$$R = f_{pc}(P) \quad (2)$$

The value of the confidence map indicates the confidence level of the CNN model regarding whether it considers the corresponding area to be a defect region. The proposed data augmentation method is constructed based on the guidance of the confidence map.

### Information dropping-based data augmentation

Information dropping is a popular data augmentation strategy that removes parts of an image to improve the perception of a CNN model with respect to originally insensitive information, thereby increasing the robustness of the model. This strategy has been widely considered to be effective in many visual scenes. Therefore, the proposed method also adopts the strategy of information dropping to perform data augmentation. Mainstream information dropping approaches are mainly divided into three categories.

1. Randomly deleting a rectangular area of an image (e.g., Cutout (DeVries & Taylor, 2017)).
2. Dividing the image into  $S \times S$  image blocks and randomly removing some image blocks (e.g., HAS (Singh & Lee, 2017)).
3. Building masks with noncontiguous structures and deleting image regions by multiplying the masks (e.g., GridMask (Chen et al., 2020)).

However, most of the above strategies, especially Cutout (DeVries & Taylor, 2017) and HAS (Singh & Lee, 2017), randomly produce invalid data augmentation results, such as those with excessive dropping and undeleted defective areas. On the one hand, excessive dropping (the red box in Fig. 5) deletes most of the context of the defect feature to change the image attributes. This results in negative optimization for the subsequent defect detection step. On the other hand, the nondeletion of defect regions (the blue box in Fig. 5) has a negligible impact on the robustness of a CNN model. Moreover, some of the above methods, such as GridMask, are based on set rules, such as discontinuous grids, which can effectively prevent excessive dropping and the nondeletion of defects. However, they cannot effectively guarantee that the deleted image content is sensitive or ignored by the CNN model.

To improve the effectiveness of information dropping for defect images, the proposed method constructs a defect dropping strategy based on the requirements of the utilized CNN model.



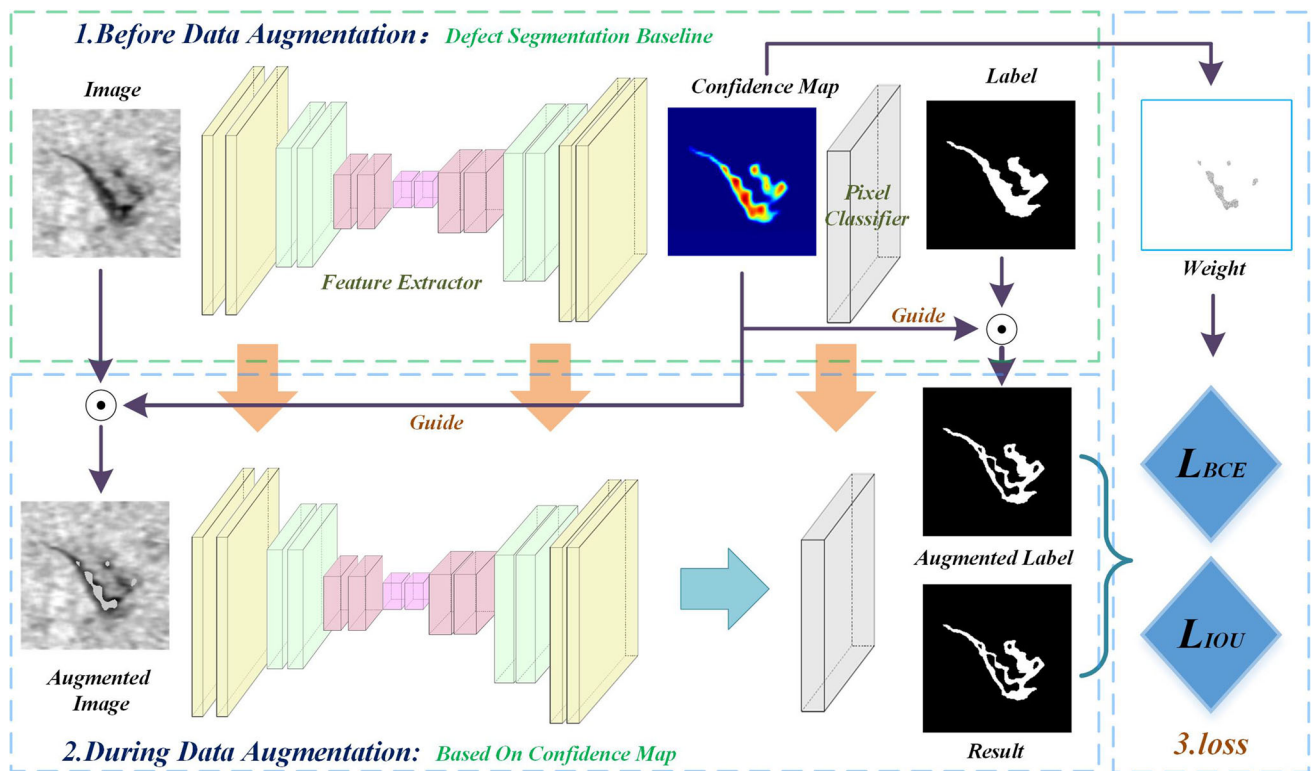


Fig. 4 The pipeline of the proposed data augmentation method

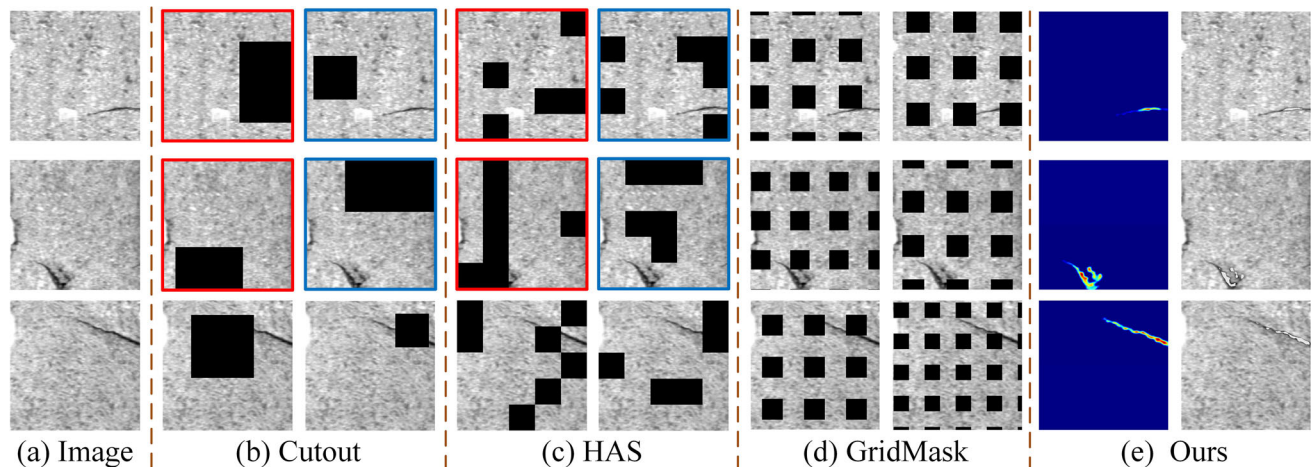


Fig. 5 A comparison of different information dropping data augmentation strategies

### Data augmentation based on CNN

Due to the inconsistencies among the characteristics of different areas within a defect, such as grayscale, texture, and position attributes, the detection difficulty levels of different defect areas are different. It is difficult for humans to accurately analyze the detection difficulty levels of different specific defect areas to set a rule for data augmentation. Since the purpose of data augmentation is to improve the ability to complete the downstream CNN tasks, the confidence map

obtained by the employed CNN model can be used as a guide for its detection capability of different defect areas. Defect areas with high confidence are considered detectable by the CNN model as easy-to-detect defect regions. Conversely, low-confidence regions can be considered difficult-to-detect defect regions.

Since the purpose of information dropping-based data augmentation is to improve a CNN model's perception of insensitive information, our augmentation scheme is to delete the defect areas that are easy to detect with the CNN model

and improve the model's attention to the difficult-to-detect defect areas. The specific steps are as follows.

1. Based on the pretraining of the CNN model, a confidence map  $P$  is obtained for each image in the training set. The closer the color of the confidence map is to red, the higher the confidence is and the easier the associated area is to detect. The confidence map can be used as a guide for data augmentation.
2. Data augmentation module  $f_d$  includes  $f_{dx}$  for input image  $X$  and  $f_{dy}$  for label  $Y$ . A threshold  $T$  is randomly chosen.

$$f_d = \langle f_{dx}, f_{dy} \rangle \quad (3)$$

For  $f_{dx}$ , the confidence values  $p_i$  are above the threshold  $T$ , and the corresponding areas in image  $X$  are occluded with the average value of the background. The regions of image  $X$  with confidence values  $p_i$  below the threshold  $T$  are retained.  $X_{aug}$  represents the augmented image

$$X_{aug} = f_{dx}(X) = \begin{cases} \text{mean}(X[Y = 0]), & p_i > T \\ X & , p_i < T \end{cases} \quad (4)$$

For  $f_{dy}$ , the areas with the label  $Y$  above  $T$  are removed and that below threshold  $T$  is retained.  $Y_{aug}$  represents the augmented label.

$$Y_{aug} = f_{dy}(Y) = \begin{cases} 0, & p_i > T \\ Y, & p_i < T \end{cases} \quad (5)$$

3. The retention rate of defect area  $\eta$  is the proportion of the retained defect area in image  $X_{aug}$  in the entire defect area of image  $X$ , defined as formula (6). We choose the range of  $\eta$  to be (0.2, 0.8).

$$\eta = \frac{X_{aug}(Y_{aug} = 1)}{X(Y = 1)} \quad (6)$$

4. Due to the long-term training with an invariant confidence map, high-confidence regions at fixed locations are removed with high probability. Therefore, the model pays more attention to low-confidence regions. However, conducting iterative training for a long time can cause overfitting, resulting in the transformation of the confidence maps of defect images. Therefore, after the CNN model iterates for a period of time, the guiding confidence map is updated according to the training progress. In this work, we set the epoch of the update cycle  $E_{uc}$  to 150 epochs as follows:

When epoch  $E! = n * 150$ , we train the segmentation

model based on the data augmentation module  $f_d$ .

$$f_{dy}(Y_{aug}) = f_{pc}(f_{fe}(f_{dx}(X_{aug}))) \quad \text{with} \\ E! = n * 150, n \in (1, 2, 3, \dots) \quad (7)$$

When epoch  $E = n * 150$ , we update the guiding confidence maps  $P$ :

$$P_{new} = f_{fe}(X) \quad \text{with} \quad E = n * 150, n \in (1, 2, 3, \dots) \quad (8)$$

The segmentation model is trained based on the new confidence map  $P_{new}$  until the next update for  $P$ .

## Loss function

To iterate and optimize the network described above, it is necessary to construct a loss function. The loss function is divided into losses for two stages: before and during data augmentation.

*Before data augmentation* The losses include the binary cross-entropy (BCE) loss and intersection-over-union (IOU) loss, which are very common and mainstream for defect segmentation.

1. The BCE loss measures the difference between the predicted class and the ground truth of each pixel, handling binary classification for each pixel.

$$Loss_{BCE}(R, Y) = -(y_i \lg r_i + (1 - y_i) \lg(1 - r_i)) \quad (9)$$

$X$  represents the predicted map, and  $Y$  represents the ground truth.  $x_i$  and  $y_i$  represent the predicted and true values of each pixel, respectively.

2. The IOU indicates the ratio between the intersection and union of  $X$  and  $Y$ . The IOU loss measures how close the IOU is to 1.

$$Loss_{IOU}(R, Y) = 1 - \frac{R \cap Y}{R \cup Y} \quad (10)$$

*During data augmentation* Since the high-confidence regions of a defect image are occluded, we set lower weights for those regions. We construct a loss weight map  $V$ . The weight is set to a random value within the range of 0.5-1 for the filled region, and the weight is set to 1 for the original region.

$$V = \begin{cases} \text{rand}(0.5, 1), & p_i > T \\ 1 & , p_i < T \end{cases} \quad (11)$$

The loss functions are constructed based on the weight map  $V$  to guide the training of the CNN model. The BCE and IOU

**Algorithm 1**

The training process of the CNN with the proposed data augmentation approach.

**Input:** Image  $X$  and its label  $Y$

**Output:** Predicted result  $R$

**Initialize** the defect segmentation network  $f_s$

Contains a feature extractor  $f_{fe}$  and a pixel classifier  $f_{pc}$

**while** epoch  $\leq 300$  (Before data augmentation):

**for**  $i=1, \dots, n$ , **do**

Obtain the predicted result  $R = f_{pc}(f_{fe}(X))$

Calculate the loss  $Loss_{total}(R, Y)$

Backpropagate the  $Loss_{total}$ , update the segmentation network  $f_s$

**end for**

**end while**

**While** epoch  $> 300$  (during data augmentation):

**If** epoch  $\% 150 = 0$ :

Obtain or update the confidence map  $P = f_{fe}(X)$

**for**  $i=1, \dots, n$ , **do**

Perform data augmentation for  $X$  to generate  $X_{aug}$  using (4)

Perform data augmentation for  $Y$  to generate  $Y_{aug}$  using (5)

Obtain the predicted result  $R = f_{pc}(f_{fe}(X_{aug}))$

Calculate the loss  $Loss_{total}(R, Y_{aug})$

Backpropagate the  $Loss_{total}$ , update the segmentation network  $f_s$

**end for**

**end while**

**Return:**  $R$

losses based on weight are shown as follows:

$$Loss_{BCE}(R, Y, V) = -v_i * (y_i \lg r_i + (1 - y_i) \lg(1 - r_i)) \quad (12)$$

$$Loss_{IOU}(R, Y, V) = 1 - \frac{(R \cdot V) \cap (Y \cdot V)}{(R \cdot V) \cup (Y \cdot V)} \quad (13)$$

The total loss is shown as follows:

$$Loss_{total} = Loss_{BCE} + \lambda Loss_{IOU} \quad (14)$$

To more clearly represent the training process of this method, the pseudocode is shown in Algorithm 1.

## Experiments

### Setup

#### Dataset

The industrial Kolektor surface defect (KSD) segmentation dataset (Tabernik et al., 2020) is chosen to verify the proposed data augmentation method. In order to simulate the situation that defects with weak intensity are difficult to collect in the industrial field, more defects with normal intensity are allocated in the training set  $D_{tr}$  and more weak defects are allocated in the test set  $D_{te}$ . Since the image size in KSD is

$500 \times 1250$ , we cut each image into blocks of size  $500 \times 500$  by using a stride of 150 pixels. The training set consists of 150 images, and the test set consists of 60 images. Figure 6 shows defect samples in the test and training sets.

### Evaluation indicators

The IOU and PA are selected as the evaluation indicators ( $EI$ ) of segmentation performance, as they are universal and widely used in different segmentation scenarios. The IOU represents the ratio between the intersection and union of the predicted results and the ground truth. PA represents the proportion of correctly classified pixels to total pixels.

$$PA = \frac{\sum_{i=0}^k P_{ii}}{\sum_{i=0}^k \sum_{j=0}^k P_{ij}} \quad (15)$$

$$IOU = \frac{\sum_{i=0}^k P_{ii}}{\sum_{j=0}^k P_{ij} + \sum_{j=0}^k P_{ji} - \sum_{i=0}^k P_{ii}} \quad (16)$$

### Implementation details

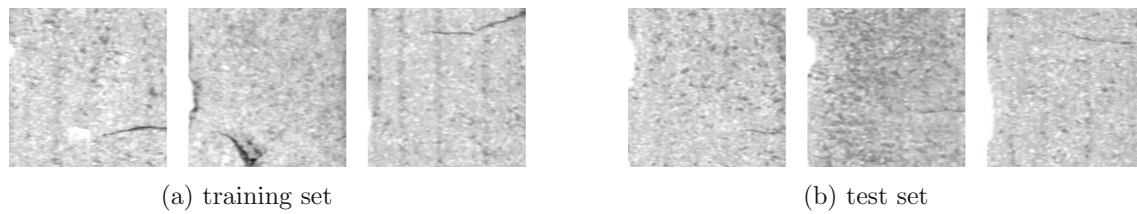
The experiment is run using an NVIDIA RTX 1080 graphics processing unit (GPU). All experiments are implemented on a deep learning framework consisting of PyTorch 1.0 and NumPy. Except for the data augmentation module, all models are set with the same parameters. The optimizer is Adam (Kingma & Ba, 2014) with a learning rate of 0.001, and the number of training epochs is 2000.

### Confidence map analysis and chosen hyperparameters

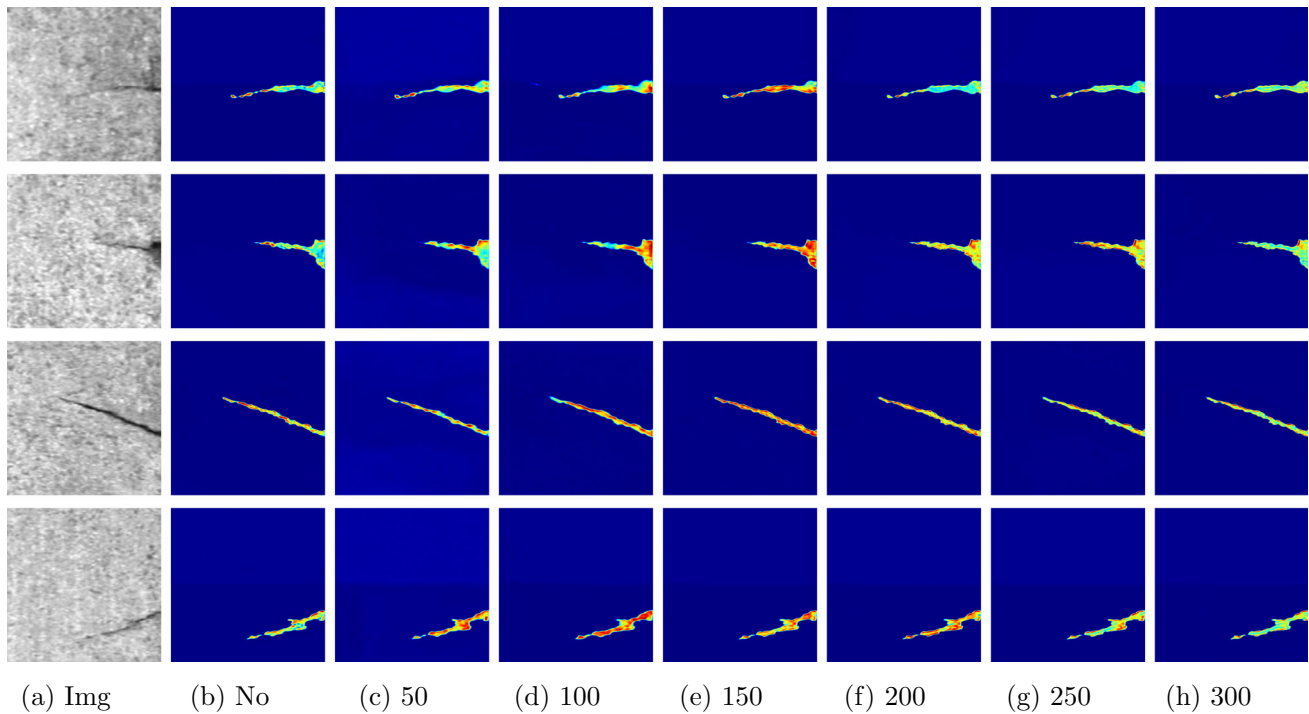
We conduct experiments to verify the effect of the proposed method on the resulting confidence maps. Figure 7 shows the confidence maps produced with and without the proposed data augmentation. We take the average value of each confidence map for the defect area as the qualitative evaluation standard, as shown in Fig. 8.

From Fig. 7, the confidence values of some defect areas are low without data augmentation in column (b), leading to bad segmentation performance. During data augmentation guided by the initial confidence maps in column (b), the model pays significantly more attention to low-confidence regions due to the higher probability of occlusion in high-confidence regions. The confidence values of the defect areas gradually increase and reach their maximum values at epochs 100–150, as shown in Fig. 8. In particular, the confidence of the regions with low confidence at the start can be significantly improved, as shown in Fig. 7. Because the confidence map exhibits greater differences between the defect areas and the background, it is able to produce better defect segmentation results during data augmentation.





**Fig. 6** The defect samples of training and test set for KSD dataset. **a** Training set. **b** Test set



**Fig. 7** The confidence maps produced without and with data augmentation at different epochs. **a** Image. **b** No. **c** 50. **d** 100. **e** 150. **f** 200. **g** 250. **h** 300

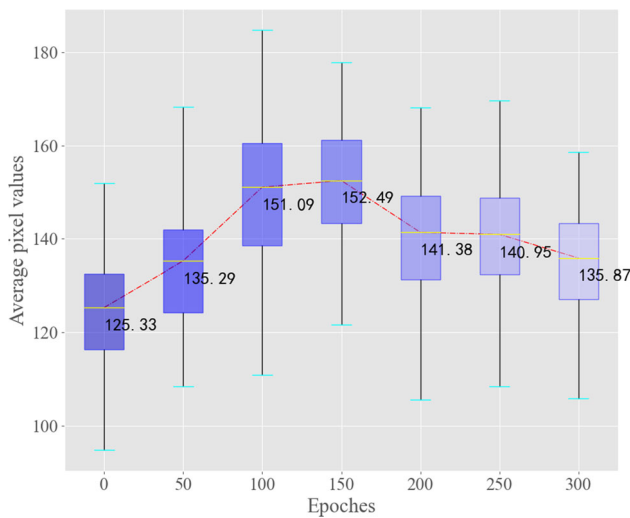
Furthermore, with the guidance confidence maps remaining unchanged, the method first significantly improves the confidence levels of all defect regions and yields their maxima within 100–150 epochs, while the subsequent training process leads to a decrease in the average confidence shown in Fig. 8. With a long training time, high-confidence regions are occluded with high probability, which makes it difficult to be noticed by CNN models. The low-confidence region is of greater concern to the model. This results in a transformation of the confidence maps. Training with the original invariant confidence map can lead to overfitting and thus decreased performance. Therefore, we update the confidence map after a period of training to prevent overfitting. According to the above experimental results, the epoch of update cycle  $E_{uc}$  in this paper is chosen as 150.

Based on the probability map, the proposed data augmentation method blocks high-probability defect areas and keeps low-probability areas. Therefore, compared to the original image  $X$ , the defect areas in the augmented image  $X_{aug}$  are

removed. The retention rate for the defect area  $\eta$  represents the ratio of the defect area of  $X_{aug}$  to  $X$ . The range of  $\eta$  is originally (0,1). However, when  $\eta$  is too high to approach 1, the defect area is not removed. When  $\eta$  is too low to approach 0, the defect areas are all deleted. Both of the above are invalid augmentations. To avoid this, we need to limit the range of  $\eta$ . We construct the experiment which set the range of  $\eta$  as (0,1), (0.2,0.8), and (0.4,0.6) separately. Table 1 is the defect segmentation performance under a different range of  $\eta$ . When the range is (0,1), it causes invalid augmentation. When the range is (0.4, 0.6),  $\eta$  is so limited that the randomness is insufficient. When the range is (0.2, 0.8), higher segmentation performance can be obtained. Therefore, we choose the range of  $\eta$  to be (0.2, 0.8) based on the experimental results.

### Comparison with state-of-the-art methods

We compare the proposed method with other state-of-the-art methods, which include Cutout (DeVries & Taylor, 2017),



**Fig. 8** The average value of the confidence map for the defect area within epochs 0–300

**Table 1** The defect segmentation performance on Unet with the different range of  $\eta$

| The range of $\eta$ | (0.4,0.6) | (0.2,0.8)    | (0,1) |
|---------------------|-----------|--------------|-------|
| IOU                 | 60.27     | <b>61.97</b> | 61.19 |
| PA                  | 90.47     | <b>99.51</b> | 99.49 |

The significance of bold in table represents the best result for each row

HAS (Singh & Lee, 2017), GridMask (Chen et al., 2020), and Cutmix (Yun et al., 2019).

1. Cutout (DeVries & Taylor, 2017): Randomly delete image blocks from the input image.
2. HAS (Singh & Lee, 2017): Divide the input image into  $n \times n$  image blocks and delete the image blocks randomly.
3. GridMask (Chen et al., 2020): Build a structured mask and multiply it with the input image.
4. Cutmix (Yun et al., 2019): Perform fusion between two input images at the block level.

**Table 2** The segmentation IOU (%) and PA (%) achieved on the KSD dataset with the proposed method and other state-of-the-art methods

| $D_{tr}$ | Seg     | El  | Baseline | Cutmix | Cutout | HAS   | GridMask | Ours         |
|----------|---------|-----|----------|--------|--------|-------|----------|--------------|
| 150      | FCN     | IOU | 50.16    | 51.59  | 57.02  | 54.10 | 57.20    | <b>59.04</b> |
|          |         | PA  | 99.35    | 99.35  | 99.43  | 99.41 | 99.44    | <b>99.45</b> |
|          | DeepLab | IOU | 52.98    | 55.46  | 57.67  | 57.65 | 57.48    | <b>58.81</b> |
|          |         | PA  | 99.38    | 99.41  | 99.43  | 99.45 | 99.44    | <b>99.47</b> |
|          | UNet    | IOU | 55.35    | 58.79  | 60.90  | 60.39 | 59.39    | <b>61.97</b> |
|          |         | PA  | 99.45    | 99.46  | 99.49  | 99.48 | 99.47    | <b>99.51</b> |
| 80       | UNet    | IOU | 54.35    | 53.15  | 55.91  | 56.08 | 55.70    | <b>57.63</b> |
|          |         | PA  | 99.40    | 99.38  | 99.40  | 99.42 | 99.41    | <b>99.45</b> |

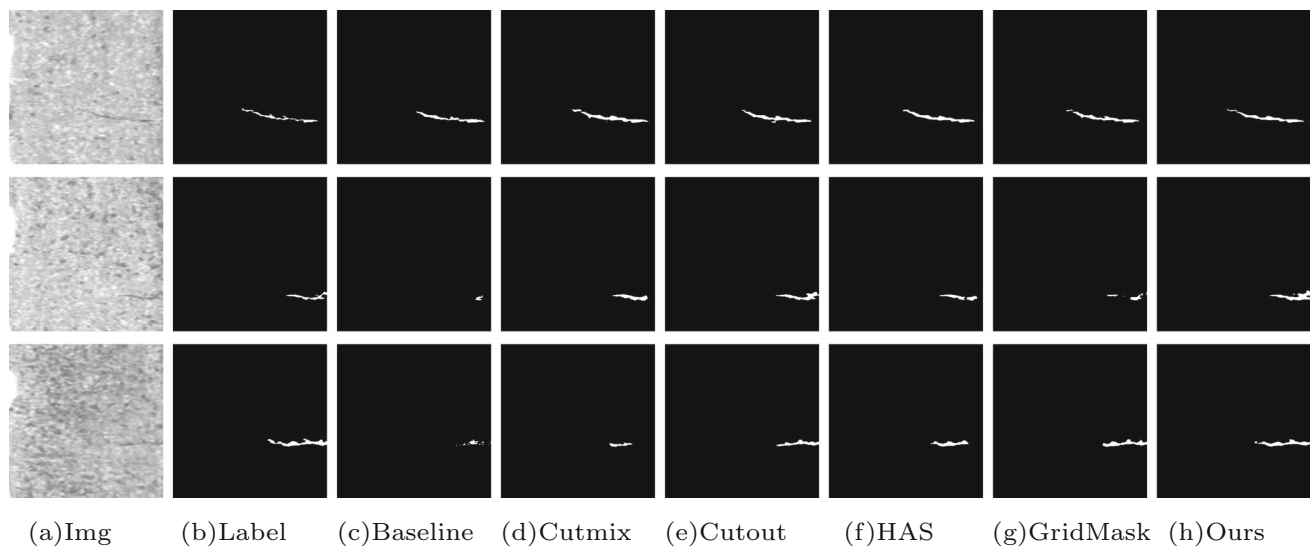
The significance of bold in table represents the best result for each row

5. The proposed method: Occlude low-confidence inputs of input images based on the confidence map obtained by downstream CNN tasks.

We conduct experiments on mainstream segmentation models, including an FCN (Long et al., 2015), DeepLab (Chen et al., 2018), and UNet (Ronneberger et al., 2015), which are used by researchers in different industrial surface defect segmentation scenarios.

1. *The FCN* (Long et al., 2015) replaces the fully connected layers in the classic classification network with convolutional layers to obtain a 2-dimensional feature map. It is followed by Softmax activation to obtain the classification information of each pixel.
2. *DeepLab* (Chen et al., 2018) adds an atrous spatial pyramid pooling (ASPP) module to the basic ResNet (He et al., 2016). ASPP involves atrous convolutions at different sampling rates and global average pooling.
3. *UNet* (Ronneberger et al., 2015) has a U-shaped network structure, which contains an encoder and a decoder. Skip connections are contained between the same stage of the encoder and decoder to ensure that the feature map incorporates more low-level features. UNet is widely used in medical image segmentation and industrial surface defect segmentation.

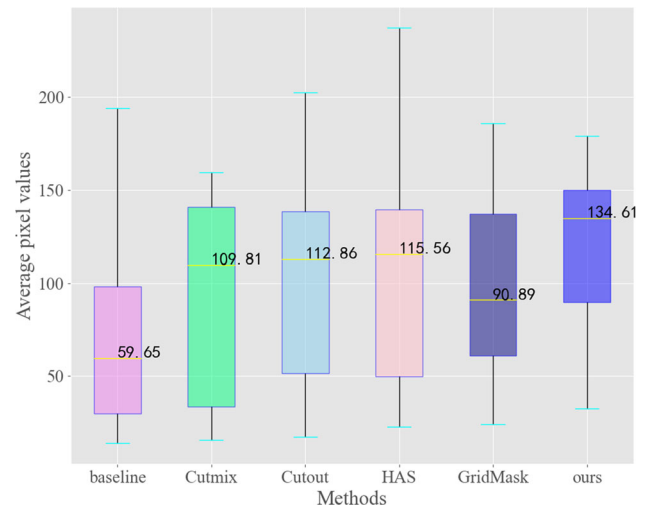
The comparison between the proposed data augmentation method and other state-of-the-art methods is shown in Table 2. Our method has significantly better accuracy than the baseline without data augmentation. Compared with other state-of-the-art methods, the proposed method also obtains higher segmentation accuracy. Moreover, the proposed data augmentation method can effectively improve the accuracy of defect segmentation when using different mainstream defect segmentation networks, such as UNet (Ronneberger et al., 2015), DeepLab (Chen et al., 2018), and an FCN (Long et al., 2015), which demonstrates the generality of our method for mainstream defect segmentation methods.



**Fig. 9** Defect segmentation results obtained without data augmentation and with various state-of-the-art data augmentation methods. **a** Image. **b** Label. **c** Baseline. **d** Cutmix. **e** Cutout. **f** HAS. **g** GridMask. **h** Ours

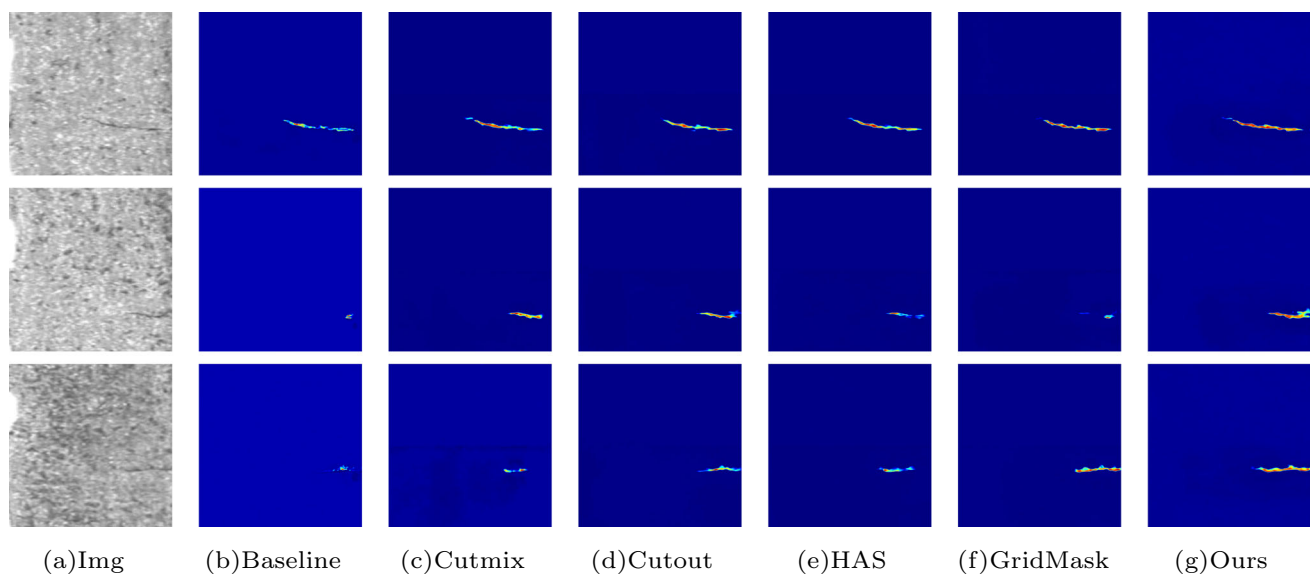
Figure 9 shows the segmentation results obtained by different data augmentation methods on UNet. For defects with normal intensities shown in the first line, the segmentation models with and without data augmentation can achieve accurate segmentation. However, for defects with weak intensities in the last two lines, the baseline cannot perform effective segmentation. All kinds of data augmentation methods can effectively improve the segmentation results of weak defects. However, other data augmentation methods, such as HAS (Singh & Lee, 2017) and GridMask (Chen et al., 2020), cannot fully and accurately identify defects of weak features. Compared to other data augmentation methods, the proposed method yields more obvious improvements for weak defects and thus obtains more accurate segmentation results.

We also analyze the confidence maps for different methods. We use the average pixel of the confidence map for the defect area as the quantitative evaluation. As shown in Fig. 10, compared with other methods, the average value obtained by the proposed method is higher, which indicates that our method can obtain a more significant confidence map. Figure 11 shows the confidence map for different methods. Specifically, although different methods can obtain similar detection results in the first line, our method can obtain a better confidence map for the defects with normal intensities, which shows that the segmentation model constructed by the proposed method has higher confidence for the defect area. Although Cutout (DeVries & Taylor, 2017) can also obtain relatively good detection results for weak feature defects in the last two lines, the proposed method can obtain more significant confidence maps than Cutout (DeVries & Taylor, 2017).

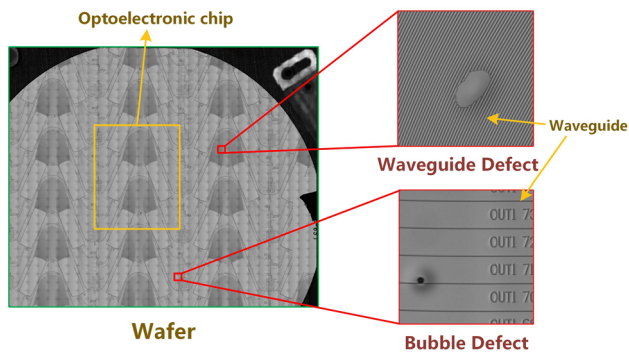


**Fig. 10** The average value of the confidence maps for different methods

The proposed data augmentation method can perform occlusion adjustment on the input image of the training set based on the probability map. Although the other methods Cutout (DeVries & Taylor, 2017), HAS (Singh & Lee, 2017) and GridMask (Chen et al., 2020) also occlude the input image, the occluded region is random, which causes invalid augmentation. The proposed method essentially increases the exposure of low-confidence regions in the model so that the model pays more attention to low-confidence regions. Moreover, the purpose of data augmentation is to improve the performance of subsequent defect segmentation. Therefore, starting from the requirements of the defect segmentation model, the proposed data augmentation is to improve the degree of attention to low-confidence regions by the downstream CNN. Therefore, the proposed method can obtain



**Fig. 11** Confidence maps for different methods. **a** Image. **b** Baseline. **c** Cutmix. **d** Cutout. **e** HAS. **f** GridMask. **g** Ours



**Fig. 12** Diagram of an optoelectronic chip and a waveguide. Waveguide defects and bubble defects appear on the wafer

more salient feature maps, including the hard-to-detect regions and thus obtain more accurate segmentation results. Moreover, difficult-to-detect regions often represent some weak feature defects, such as weak contrast defects. Therefore, the proposed method can significantly improve the detection ability of defects with weak features.

In order to verify data augmentation in the case of different data volume levels, we reduce the training set  $D_{tr}$  from 150 to 80. As shown in Table 2, compared with other methods, the proposed method can further significantly increase the defect segmentation accuracy over the baseline with the smaller training set. This shows the effectiveness of our method under different data volumes.

### An application of the proposed method

The proposed method is applied to an actual optoelectronic chip production line. An optoelectronic chip is an important component of optical communication equipment. As shown

in Fig. 12, a large number of waveguide lines are arranged on the optoelectronic chips of a wafer for the propagation of optical information. During the manufacturing process, due to the small sizes and dense arrangement of the optical waveguides, waveguide defects (WDs), such as breakages or crossovers, may occur, affecting the efficiency and accuracy of the optical wave transmission process. Moreover, the packaging of the optoelectronic chips may cause bubble defects (BDs). We construct an optical chip defect dataset named OCDD, which contains a total of 317 BD images and 224 WD images. We randomly divide the training set and test set at a ratio of 2:1.

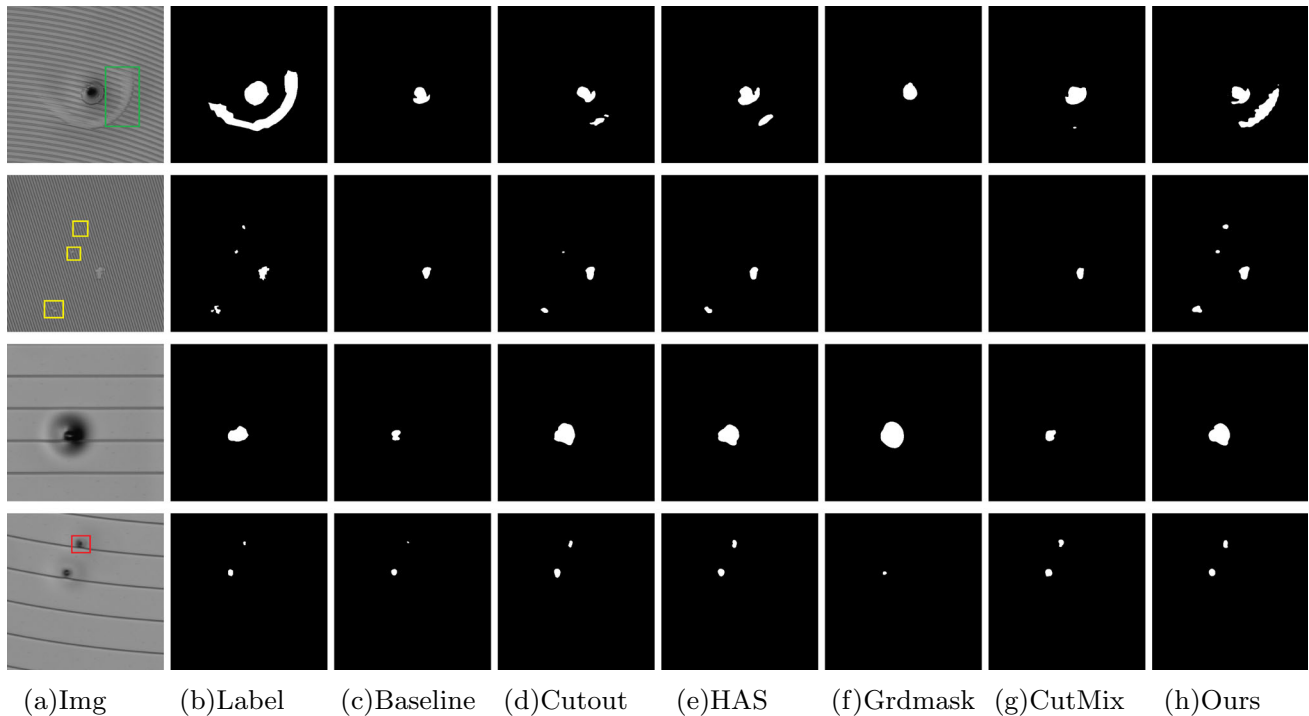
We construct comparative experiments for OCDD with the proposed method and other state-of-the-art methods. Based on the results in “[Comparison with state-of-the-art methods](#)” section, we employ UNet (Ronneberger et al., 2015) and the proposed data augmentation method to construct a segmentation model. The baseline is UNet (Ronneberger et al., 2015) without any data augmentation. As shown in Table 3, the proposed data augmentation approach significantly improves the accuracy of defect segmentation and achieves higher accuracy than the other methods. Furthermore, as shown in Fig. 13, the proposed data augmentation method can effectively detect the WDs with weak features, such as those with similar background colors (green boxes) and small areas (yellow boxes), which cannot be effectively detected by the baseline without data augmentation and other methods. The proposed data augmentation method can also detect BDs (red boxes) more completely than the baseline, which is very important for judging the severity of defects. Since the optoelectronic chip detection accuracy is limited by the defects with weak features and a small area, which are of greater concern to the



**Table 3** The defect segmentation performance of the proposed methods and other state-of-the-art methods on an optoelectronic chip

| Method | <i>EI</i> | Baseline | Cutout | HAS          | GirdMask | Cutmix | Ours         |
|--------|-----------|----------|--------|--------------|----------|--------|--------------|
| WD     | IOU       | 51.05    | 56.86  | 56.32        | 52.63    | 55.47  | <b>58.38</b> |
|        | PA        | 99.41    | 99.46  | 99.46        | 99.43    | 99.44  | <b>99.48</b> |
| BD     | IOU       | 69.89    | 66.17  | 69.10        | 66.23    | 69.30  | <b>69.90</b> |
|        | PA        | 98.82    | 99.30  | <b>99.37</b> | 99.30    | 99.17  | 99.28        |
| Total  | IOU       | 61.75    | 62.13  | 63.95        | 60.96    | 63.87  | <b>64.91</b> |
|        | PA        | 99.08    | 99.32  | <b>99.40</b> | 99.35    | 99.29  | 99.37        |

The significance of bold in table represents the best result for each row

**Fig. 13** Defect segmentation results of the optoelectronic chip obtained without data augmentation and with our data augmentation method. **a** Image. **b** Label. **c** Baseline. **d** Cutout. **e** HAS. **f** GridMask. **g** CutMix. **h** Ours

engineers, the proposed method can be effectively applied in optoelectronic chip defect detection scenarios.

## Discussion

The information density of different image features is different. For example, defect features that are frequent and easy to detect tend to have a lower density, whereas rare and difficult-to-detect defect features tend to have higher density. In this paper, high-confidence defect regions have lower information density, whereas low-confidence defect regions have higher information density for CNN. During the CNN training process, a large amount of redundant feature information, such as similar defect features, occupies the model's attention and causes model degradation (Lin et al., 2017); this is called the easy sample overwhelming problem.

The essence of data augmentation is the resampling and combination of existing feature information. Therefore, more attention should be paid to feature information with higher density during resampling for data augmentation. As mentioned above, mainstream data augmentation methods, such as GANs and Cutout (DeVries & Taylor, 2017), are random and do not focus on defective samples with higher information density. Moreover, the GAN training process itself has a mode collapse problem, which tends to generate repeated samples with low density, contrary to the original intention of data augmentation. The main strategy of the proposed data augmentation approach is to remove low-density areas to increase the attention of the CNN model to high-density regions. This technique is essentially based on the idea of hard sample mining, which is performed at the pixel level instead of at the image level.

## Conclusion

In this paper, we propose a simple plug-and-play defect image augmentation method based on CNN requirements, which can be directly applied to existing defect segmentation methods. The proposed method utilizes confidence maps to remove the high-confidence regions of an image to suppress the attention to easy-to-detect areas and increase the attention to difficult-to-detect areas. We construct adequate quantitative and qualitative experiments on metal surfaces and optical chips to compare our method with state-of-the-art data augmentation approaches. The experimental results show that the proposed method can effectively achieve improved segmentation performance, especially for defects with weak features, and is suitable for different mainstream defect segmentation models.

In the future, some technical issues can still be further improved, and the proposed method can be extended. (1) This paper is based on the confidence map obtained from the segmentation model to guide data augmentation. In the follow-up, we can further design specific performance indicators to guide data augmentation according to other needs of the industrial site. (2) We can further apply the idea of performing data augmentation based on CNN requirements to various industrial vision tasks, such as defect classification and defect object detection.

**Acknowledgements** This work was supported in part by the National Key R&D Program of China under Grant 2018YFB1700500 and the Key Research and Development Program of HubeiChina (2020BAB106).

## References

- Bissoto, A., Valle, E., & Avila, S. (2021). Gan-based data augmentation and anonymization for skin-lesion analysis: A critical review. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 1847–1856).
- Chen, J. N., Sun, S., He, J., Torr, P. H., Yuille, A., & Bai, S. (2021). Transmix: Attend to mix for vision transformers. arXiv preprint [arXiv:2111.09833](https://arxiv.org/abs/2111.09833).
- Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. (2018). Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(4), 834–848.
- Chen, P., Liu, S., Zhao, H., & Jia, J. (2020). Gridmask data augmentation. arXiv preprint [arXiv:2001.04086](https://arxiv.org/abs/2001.04086).
- Cheng, K.C.-C., Chen, L.L.-Y., Li, J.-W., Li, K.S.-M., Tsai, N.C.-Y., Wang, S.-J., et al. (2021). Machine learning-based detection method for wafer test induced defects. *IEEE Transactions on Semiconductor Manufacturing*, 34(2), 161–167.
- Choi, J., Kim, T., & Kim, C. (2019). Self-ensembling with gan-based data augmentation for domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6830–6840).
- Choi, Y., Choi, M., Kim, M., Ha, J.-W., Kim, S., & Choo, J. (2018). Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *2018 IEEE/CVF conference on computer vision and pattern recognition* (pp. 8789–8797).
- Chuanfei, H., & Wang, Y. (2020). An efficient convolutional neural network model based on object-level attention mechanism for casting defect detection on radiography images. *IEEE Transactions on Industrial Electronics*, 67(12), 10922–10930.
- DeVries, T., & Taylor, G. W. (2017). Improved regularization of convolutional neural networks with cutout. arXiv preprint [arXiv:1708.04552](https://arxiv.org/abs/1708.04552).
- Dong, H., Kechen Song, Yu., He, J. X., Yan, Y., & Meng, Q. (2020). Pga-net: Pyramid feature fusion and global context attention network for automated surface defect detection. *IEEE Transactions on Industrial Informatics*, 16(12), 7448–7458.
- Fan, R., Wang, H., Cai, P., Jin, W., Bocus, J., Qiao, L., & Liu, M. (2021). Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation. *IEEE/ASME Transactions on Mechatronics*.
- Ghiassi, G., Cui, Y., Srinivas, A., Qian, R., Lin, T.-Y., Cubuk, E. D., Le, Q. V., & Zoph, B. (2021). Simple copy-paste is a strong data augmentation method for instance segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2918–2928).
- Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial networks. arXiv preprint [arXiv:1406.2661](https://arxiv.org/abs/1406.2661).
- He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).
- Hendrycks, D., Mu, N., Cubuk, E. D., Zoph, B., Gilmer, J., & Lakshminarayanan, B. (2019). Augmix: A simple data processing method to improve robustness and uncertainty. arXiv preprint [arXiv:1912.02781](https://arxiv.org/abs/1912.02781).
- Hsu, C.-Y., & Liu, W.-C. (2021). Multiple time-series convolutional neural network for fault detection and diagnosis and empirical study in semiconductor manufacturing. *Journal of Intelligent Manufacturing*, 32(3), 823–836.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700–4708).
- Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 5967–5976).
- Jain, S., Seth, G., Paruthi, A., Soni, U., & Kumar, G. (2020). Synthetic data augmentation for surface defect detection and classification using deep learning. *Journal of Intelligent Manufacturing*, 1–14.
- Kim, Y., Cho, D., & Lee, J.-H. (2021). Wafer defect pattern classification with detecting out-of-distribution. *Microelectronics Reliability*, 122, 114157.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25.
- Le, X., Mei, J., Zhang, H., Zhou, B., & Xi, J. (2020). A learning-based approach for surface defect detection using small image datasets. *Neurocomputing*, 408, 112–120.
- Li, W., Chen, J., Cao, J., Ma, C., Wang, J., Cui, X., & Chen, P. (2022). Eid-gan: Generative adversarial nets for extremely imbalanced data augmentation. *IEEE Transactions on Industrial Informatics*.
- Lin, C.-T., Huang, S.-W., Yen-Yi, W., & Lai, S.-H. (2020). Gan-based day-to-night image style transfer for nighttime vehicle detection.

- IEEE Transactions on Intelligent Transportation Systems*, 22(2), 951–963.
- Lin, T. Y., Goyal, P., Girshick, R., He, K., & Doll, P. (2017). Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision* (pp. 2980–2988).
- Liu, C., Wang, K., Wang, Y., & Yuan, X. (2021). Learning deep manifold structure feature representation for quality prediction with an industrial application. *IEEE Transactions on Industrial Informatics*.
- Liu, J., Guo, F., Gao, H., Li, M., Zhang, Y., & Zhou, H. (2021). Defect detection of injection molding products on small datasets using transfer learning. *Journal of Manufacturing Processes*, 70, 400–413.
- Liu, J., Wang, C., Hai, S., Bo, D., & Tao, D. (2019). Multistage gan for fabric defect detection. *IEEE Transactions on Image Processing*, 29, 3388–3400.
- Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3431–3440).
- Luo, M., Cao, J., Ma, X., Zhang, X., & He, R. (2021). Fa-gan: face augmentation gan for deformation-invariant face recognition. *IEEE Transactions on Information Forensics and Security*, 16, 2341–2355.
- Müller, S. G., & Hutter, F. (2021). Trivialaugment: Tuning-free yet state-of-the-art data augmentation. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 774–782).
- Naghizadeh, A., Xu, H., Mohamed, M., Metaxas, D. N., & Liu, D. (2021). Semantic aware data augmentation for cell nuclei microscopical images with artificial neural networks. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 3952–3961).
- Niu, S., Li, B., Wang, X., & Lin, H. (2020). Defect image sample generation with gan for improving defect recognition. *IEEE Transactions on Automation Science and Engineering*, 17(3), 1611–1622.
- Pan, L., Rogulin, R., & Kondrashev, S. (2021). Artificial neural network for defect detection in ct images of wood. *Computers and Electronics in Agriculture*, 187, 106312.
- Ren, X., Lin, W., Yang, X., Yu, X., & Gao, H. (2022). Data augmentation in defect detection of sanitary ceramics in small and non-iid datasets. *IEEE Transactions on Neural Networks and Learning Systems*.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *International conference on medical image computing and computer-assisted intervention* (pp. 234–241). Springer.
- Sassi, P., Tripicchio, P., & Avizzano, C. A. (2019). A smart monitoring system for automatic welding defect detection. *IEEE Transactions on Industrial Electronics*, 66(12), 9641–9650.
- Schlosser, T., Friedrich, M., Beuth, F., & Koworko, D. (2022). Improving automated visual fault inspection for semiconductor manufacturing using a hybrid multistage system of deep neural networks. *Journal of Intelligent Manufacturing*, 33(4), 1099–1123.
- Singh, K. K., & Lee, Y. J. (2017). Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization. In *2017 IEEE international conference on computer vision (ICCV)* (pp. 3544–3553). IEEE.
- Su, B., Chen, H., Chen, P., Bian, G.-B., Liu, K., & Liu, W. (2021). Deep learning-based solar-cell manufacturing defect detection with complementary attention network. *IEEE Transactions on Industrial Informatics*, 17(6), 4084–4095.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2818–2826).
- Tabernik, D., Šela, S., Skvarč, J., & Skočaj, D. (2020). Segmentation-based deep-learning approach for surface-defect detection. *Journal of Intelligent Manufacturing*, 31(3), 759–776.
- Tang, J., Zhou, H., Wang, T., Jin, Z., Wang, Y., & Wang, X. (2022). Cascaded foreign object detection in manufacturing processes using convolutional neural networks and synthetic data generation methodology. *Journal of Intelligent Manufacturing* 1–17.
- Uricar, M., Sistu, G., Rashed, H., Vobecky, A., Kumar, V.R., Krizek, P., Burger, F., & Yogamani, S. (2021). Let's get dirty: Gan based data augmentation for camera lens soiling detection in autonomous driving. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 766–775).
- Wang, X., Man, Z., You, M., & Shen, C. (2017). Adversarial generation of training examples: applications to moving vehicle license plate recognition. arXiv preprint [arXiv:1707.03124](https://arxiv.org/abs/1707.03124).
- Xuan, Q., Chen, Z., Liu, Y., Huang, H., Bao, G., & Zhang, D. (2018). Multiview generative adversarial network and its application in pearl classification. *IEEE Transactions on Industrial Electronics*, 66(10), 8244–8252.
- Yang, B., Liu, Z., Duan, G., & Tan, J. (2021). Mask2defect: A prior knowledge based data augmentation method for metal surface defect inspection. *IEEE Transactions on Industrial Informatics*.
- Yun, J. P., Shin, W. C., Koo, G., Kim, M. S., Lee, C., & Lee, S. J. (2020). Automated defect inspection system for metal surfaces based on deep learning and data augmentation. *Journal of Manufacturing Systems*, 55, 317–324.
- Yun, S., Han, D., Oh, S. J., Chun, S., Choe, J., & Yoo, Y. (2019). Cutmix: Regularization strategy to train strong classifiers with localizable features. In *Proceedings of the IEEE/CVF international conference on computer vision* (pp. 6023–6032).
- Zhang, H., Cisse, M., Dauphin, Y. N., & Lopez-Paz, D. (2017). mixup: Beyond empirical risk minimization. arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412).
- Zhang, L., Wang, P., Li, H., Li, Z., Shen, C., & Zhang, Y. (2020). A robust attentional framework for license plate recognition in the wild. *IEEE Transactions on Intelligent Transportation Systems*, 22(11), 6967–6976.
- Zheng, Z., Yang, X., Yu, Z., Zheng, L., Yang, Y., & Kautz, J. (2019). Joint discriminative and generative learning for person re-identification. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 2138–2147).
- Zheng, Z., Zheng, L., & Yang, Y. (2017). Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In *Proceedings of the IEEE international conference on computer vision* (pp. 3754–3762).
- Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223–2232).

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.