

Article

# Anomaly Detection in the Production Process of Stamping Progressive Dies Using the Shape- and Size-Adaptive Descriptors

Liang Ma and Fanwu Meng \* 

School of Mechanical Engineering, Beijing Institute of Technology, Beijing100081, China; liangma@bit.edu.cn

\* Correspondence: mfanwu@bit.edu.cn

**Abstract:** In the production process of progressive die stamping, anomaly detection is essential for ensuring the safety of expensive dies and the continuous stability of the production process. Early monitoring processes involve manually inspecting the quality of post-production products to infer whether there are anomalies in the production process, or using some sensors to monitor some state signals during the production process. However, the former is an extremely tedious and time-consuming task, and the latter cannot provide warnings before anomalies occur. Both methods can only detect anomalies after they have occurred, which usually means that damage to the die has already been caused. In this paper, we propose a machine-vision-based method for real-time anomaly detection in the production of progressive die stamping. This method can detect anomalies before they cause actual damage to the mold, thereby stopping the machine and protecting the mold and machine. In the proposed method, a whole continuous motion scene cycle is decomposed into a standard background template library, and the potential anomaly regions in the image to be detected are determined according to the difference from the background template library. Finally, the shape- and size-adaptive descriptors of these regions and corresponding reference regions are extracted and compared to determine the actual anomaly regions. The experimental results demonstrate that this method not only achieves satisfactory accuracy in anomaly detection during the production of progressive die stamping, but also attains competitive performance levels when compared with methods based on deep learning. Furthermore, it requires simpler preliminary preparations and does not necessitate the adoption of the deep learning paradigm.



**Citation:** Ma, L.; Meng, F. Anomaly Detection in the Production Process of Stamping Progressive Dies Using the Shape- and Size-Adaptive Descriptors. *Sensors* **2023**, *23*, 8904. <https://doi.org/10.3390/s23218904>

Academic Editor: Marco Leo

Received: 23 September 2023

Revised: 29 October 2023

Accepted: 31 October 2023

Published: 1 November 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Keywords:** stamping progressive die; anomaly detection; shape- and size-adaptive descriptor; machine vision

## 1. Introduction

The progressive die is a stamping device that efficiently produces parts through continuous stamping of metal sheets, utilizing a press machine and a mold, based on the deformation theory of metal thin plates [1,2]. Smaller sheet metal parts that are needed in large quantities are typically manufactured using progressive dies due to the process's stability, high production rate, and automation. Progressive dies have multiple stations, each performing one or more stamping operations [3]. The multi-station stamping progressive die is a type of advanced and efficient processing equipment for forming sheet metal parts, which can complete punching, bending, forming, and other stamping processes, and is widely used in modern industrial production. A characteristic of the production process of stamping progressive dies is periodic motion, which ensures highly stable production quality. However, the state of the progressive die often includes anomalies such as residual processing waste and foreign object splashes on the production line, as well as contamination and severe deformation of the workpiece [4,5]. If these anomalies occur at the stations of the progressive die and stamping continues, these anomalies may cause damage to the expensive mold and even pose a threat to production safety. Therefore,

real-time monitoring of these stations and pausing the machine when anomalies occur at the stations are essential for protecting the mold and ensuring the normal operation of the processing process.

The early traditional approach involved workers constantly monitoring the products being manufactured on the production line. Upon detecting any anomalies in the products, the workers would immediately halt production and inspect the machinery. However, this was an extremely tedious and time-consuming process. Furthermore, the workers' attention could potentially be diverted, leading to a delay in the detection of any abnormalities. To improve quality, state-of-the-art sensors are being used to replace visual inspections [6]. More recent methods for stamping process monitoring are based on the analysis of status signals derived from sensors installed on the processing equipment, which use the tonnage signature, acoustic signature, vibration signature, pressure signature, thermal signature, and other signatures as health indicators to determine the working condition of stamping progressive dies. Sah and Mahayotsanun et al. [7,8] used an array of tooling-integrated force sensors to measure contact pressure distribution across the sheet metal tooling interface for stamping process monitoring. Xu et al. [9] combined sensing techniques and the hidden Markov model to develop a fault diagnosis system, which enables adaptability and flexibility in monitoring industrial manufacturing processes. Li et al. [10] proposed an audio signal processing approach to inspect manufacturing equipment for tool wear. Kim et al. [11] integrated the principal component analysis technique and tonnage sensing system to perform stamping process inspection. These methods can identify an abnormal status in stamping equipment when malfunctions happen and, to some extent, avoid greater economic losses. Unfortunately, the biggest obstacle in the actual production process is the inability to monitor anomalies in real time, that is, to detect anomalies when they occur but have not yet caused a failure, and to immediately stop the operation of the machine. Currently, signal changes can only be detected when a failure has already occurred, which does not allow for taking measures in advance to limit the failure of processing equipment.

Machine vision technology can be employed for inspection purposes. Its fundamental principle involves the use of industrial cameras to continuously capture images of the target area, also known as the Region of Interest (ROI). Suitable algorithms are then applied to analyze the captured images to ascertain whether the target meets the requirements. Methods based on machine vision can achieve low-cost, high-precision, real-time inspection of target objects, without exerting any external influence on the production process [12]. Traditional computer vision techniques were often employed to detect surface defects in early studies. Ghorai et al. [13] employed wavelet features combined with a support vector machine to localize steel surface defects. Xie et al. [14] proposed an approach based on data augmentation and a support vector machine to detect defect patterns in noisy images. Liu et al. [15] proposed a model to project the local texture distribution into the low-dimension space, and an adaptive threshold was chosen to distinguish defects from the background. Truong and Kim [16] improved Otsu's method via an entropy weighting scheme to segment small defect regions. Substantial research has been conducted on vision-based manufacturing process monitoring approaches. Martinez et al. [17] compared the information extracted from an industrial camera placed on top of a steel framing machine prototype with the manufacturing information available from the building information model to perform the pre-inspection of steel frame manufacturing. Lin [18] introduced a new adaptive vision-based method combining discrete wavelet transform-based feature extraction and support vector machine classification for automated inspection in manufacturing. Liu et al. [19] developed a product quality classifier based on a sparse multikernel least squares support vector machine to enable the supervision of assembly production lines. Gamage et al. [20] investigated possible defect detection methodologies and subsequently proposed a system capable of the real-time monitoring of defects in the cast extrusion manufacturing process.

So far, few vision-based methods have been proposed for online stamping process monitoring. The vision-based detection techniques related to stamping progressive dies mainly focus on offline workpiece quality monitoring, such as threshold-based methods [21,22],

edge-based methods [23], and template matching-based methods [24], and these methods were compared in [25]. Stamping workpiece quality monitoring methods are mainly used for offline defect detection.

Although automatic surface defect detection via computer vision techniques has shown good performance in detecting specific surface defects, these methods could be further improved, since the complex feature extraction methods are often carefully designed based on the human experience.

In contrast, deep-learning-based automatic feature extraction methods have a strong pattern recognition ability, without requiring manual extraction of features. Researchers applied deep learning to surface defect detection, achieving greater accuracy than conventional methods. Networks such as VGG [26], GoogLeNet [27], and ResNet [28], which achieved high accuracy in natural image classification, have been applied to industrial images for classification [29–32] or feature extraction [33,34]. As a result, deep-learning-based defect detection methods have gained increasing popularity with applications in various industrial settings. Supervised methods are usually preferred when diversified and adequate defective samples can be easily collected and labeled. Yin [35] utilized Yolo V3 to detect damage defects in sewage pipelines and obtained 85.37% mean average precision (mAP). Feng [36] proposed an improved encoding–decoding network based on feature image fusion to detect cracks in hydroelectric dam images acquired by an unmanned aerial vehicle (UAV). Xiao [37] introduced a hierarchical-feature-based convolution neural network (H-CNN) model to detect oil leaks in freight trains. Due to the high level of standardization in industrial processes, instances of labeled damage patterns are seldom available. Infrequent deviations from normal conditions make it extremely challenging to gather an adequate number of labeled examples that accurately depict representative types of defects [38]. Manual delineation of the rectangular frame, as well as pixel-by-pixel segmentation, requires significant effort and assets, making collecting numerous defective samples and covering all defect types strenuous. A self-supervised learning strategy is capable of addressing these issues. Detone et al. [39] introduced a self-supervised framework for training interest point detectors that are applicable to multi-view geometry problems. In this framework, a homographic adaptation approach is proposed to generate pseudo-ground-truth interest points for self-supervised training. Araslanov and Roth [40] devised a data augmentation technique within a self-supervised framework that is trained on co-evolving pseudo labels, eliminating the need for cumbersome additional training rounds. Pautrat et al. [41] further expanded the self-supervised learning method in [39] to detect line segments. Xu [42] proposed SEDD, where a self-supervised learning strategy is utilized to address the scarcity of defective samples. Tasi et al. [43] proposed a reconstruction model based on convolutional autoencoders for the rapid and reliable detection of defects. These defects are trained using unsupervised learning strategies, which classify test images as defective or flawless but are unable to achieve pixel-level defect detection. Chow et al. [44] achieved good results in detecting concrete defects through the use of convolutional autoencoders to detect defects in concrete structures. Sean Givnan and colleagues applied autoencoders for anomaly detection in industrial motors [45]. Mishra et al. [46] proposed a novel transformer-based anomaly detection method that combines reconstruction-based methods with patch embedding. Wu et al. [47] proposed a self-supervised framework for comparison and recovery, which aims to learn generalized representations from unmarked defect images and improve the performance of various defect detection methods.

Dynamic motion scene abnormality monitoring methods are more suitable for stamping process monitoring. At present, dynamic scene modeling is used to detect moving objects in complex motion scenes. Common feature dynamic scene modeling methods include the hybrid Gaussian modeling algorithm, the Bayesian background modeling algorithm, the non-parametric kernel density estimation method, and the ViBe scenario modeling method [48–50]. However, the aforementioned modeling methods are not applicable when the entire Region of Interest (ROI) is in motion. This is because they all utilize the principle of residuals to subtract the pre-established background from the captured

image in order to identify the moving regions. Consequently, they cannot detect deviations caused by loose parts in the device, nor can they protect the processing equipment with high accuracy.

In summary, the following issues exist in the monitoring process of the progressive die stamping production: (1) most equipment undergoes cyclical rather than static changes; (2) anomalies are sporadic and non-prior; (3) the workpiece will produce elastic deformation image differences; and (4) the pre-training of methods based on deep learning is time-consuming. To solve these problems, a shape- and size-adaptive descriptor (SSAD) is constructed, which is robust to all possible types of interference, to ensure high detection accuracy and thus ensure the normal operation of the machine.

The organization of this article is as follows. In Section 2, the methods applied to anomaly detection in the stamping process are described. Section 3 describes the comparative experiments conducted to verify the effectiveness of this method, where we compare our method with another descriptor and several popular deep-learning-based anomaly detection methods. The conclusions are described in Section 4.

## 2. Materials and Methods

Considering that the region of stamping production lines to be inspected is a periodic motion scene and the occurrence of anomalies is contingent and non-transcendental, a novel method for detecting anomalies in periodic scenes was developed in this study. The method consists of three main steps: (1) image segmentation, (2) SSAD construction, and (3) t-distribution-function-based anomaly region determination. This method was briefly introduced in our earlier work [51], and this paper will provide a more detailed description of our approach.

### 2.1. Image Segmentation

The first stage of image segmentation involves building the standard template image library by decomposing the periodic motion scene of the stamping production line, using the method proposed by Wang et al. [52].

Assuming that the camera position and shooting angle are fixed, a continuous series of images of the station captured at continuous variable time  $t$ ,  $Im(x, y, t)$ , is defined as its continuous motion scene. As suggested by Cutler and Davis [53], if the continuous motion scene of a station  $Im(x, y, t)$  satisfies the equation

$$Im(x, y, t) = Im(x, y, t + P), (P > 0), \quad (1)$$

where  $P$  is a constant period, then the continuous motion scene is defined as a periodic motion scene.

To decompose the continuous motion scene into a discrete series of images, it is sampled with a constant sampling period  $T$ . If  $T$  satisfies the Shannon sampling theorem, the continuous motion scene information can be completely saved in a discrete series of images. The procedure for this can be mathematically described as multiplying  $Im(x, y, t)$  by the sampling function  $\delta_T(t)$  and then integrating the resulting product with respect to  $t$ :

$$Im(x, y, nT) = \int_0^{+\infty} Im(x, y, t) \delta_T(t - nT) dt, \quad (n = 0, \pm 1, \pm 2 \dots). \quad (2)$$

When a stamping cycle is sampled,  $\frac{P}{T}$  images of the stamping production line,  $Im(x, y, nT)$  ( $n = 1, 2 \dots \frac{P}{T}$ , and in the following,  $Im(x, y, nT)$  is marked as  $Im_n$ ), are obtained. These images are defined as the standard template image library and can approximately replace  $Im(x, y, t)$ . Once the standard template image library has been constructed, the next step is to select the best matching image for the image to be detected,  $Im_{wt}$ , which is captured during production and is used to inspect the stamping production line for anomalies.

A clustering strategy is utilized to accelerate the matching step. It starts from the image with the smallest subscript picked among standard template images that do not yet fall within a cluster and then calculates the similarity measure,  $Sim$ , between this image and the image with the smallest subscript in the last cluster.  $Sim$  can be computed using the following formula:

$$Sim = \frac{1}{1 + \sum_{x=1}^M \sum_{y=1}^N |Im_1(x,y) - Im_2(x,y)|} \tag{3}$$

where  $Im_1$  and  $Im_2$  represent images, and  $M$  and  $N$  are the size of the image. If the similarity measure is larger than a given threshold, this image will be assigned to the last cluster. Otherwise, we build a new cluster with it. We proceed this way until every image belongs to a cluster. By first computing the similarity measure between the image  $Im_{wt}$  and the images with the smallest subscript of every cluster, we can find the cluster most similar to  $Im_{wt}$ . We further calculate the similarity measure between  $Im_{wt}$  and every standard template image in this cluster and define the image that causes the similarity measure to take the maxima as the best matching image of  $Im_{wt}$ .

The similarity curve of the periodic motion scene  $Im(x,y,t)$  and the image  $Im_{wt}$  is shown in Figure 1.

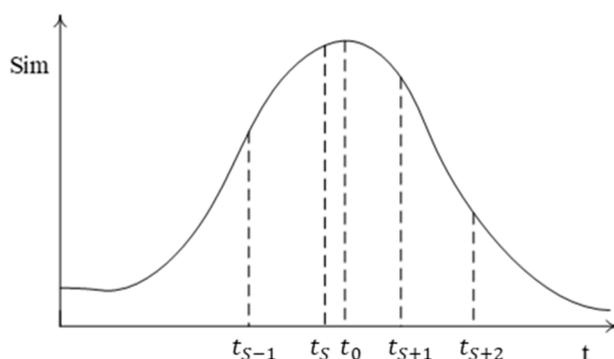


Figure 1.  $Sim-t$  curve.

The curve takes the maximum value at  $t_0$ , and  $t_s = t_{s-1} + T$ ,  $t_{s+1} = t_s + T$ . However, the template image corresponding to  $t_0$  is often not available because the standard template image library is a discrete series of images. In practice, the image  $Im_s$  corresponding to  $t_s$  that is closest to  $t_0$  is determined as the best matching image for  $Im_{wt}$ . A time interval  $t_g$  ( $0 \leq t_g < T$ ) exists between the image  $Im_{wt}$  and standard template image  $Im_s$ , in which a certain translation occurs between the two images in the spatial domain.

The image pair is calculated by matching the image  $Im_{wt}$  to the standard template image library. Thereafter, image registration is carried out to align this image pair. The geometric transformation model between the standard template image  $Im_s$  and image  $Im_{wt}$  can be approximated as the affine transformation model:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \tag{4}$$

Considering that the feature-based image registration method offers relatively high accuracy and efficiency [54] and that SURF features exhibit certain robustness to noise and affine transformation [55,56], the features of images  $Im_s$  and  $Im_{wt}$  are extracted using the method developed by Bay et al. [55]. Thereafter, by means of feature matching, calculation of the affine transformation model parameters, image transformation, and resampling steps, the image pair is registered [57].

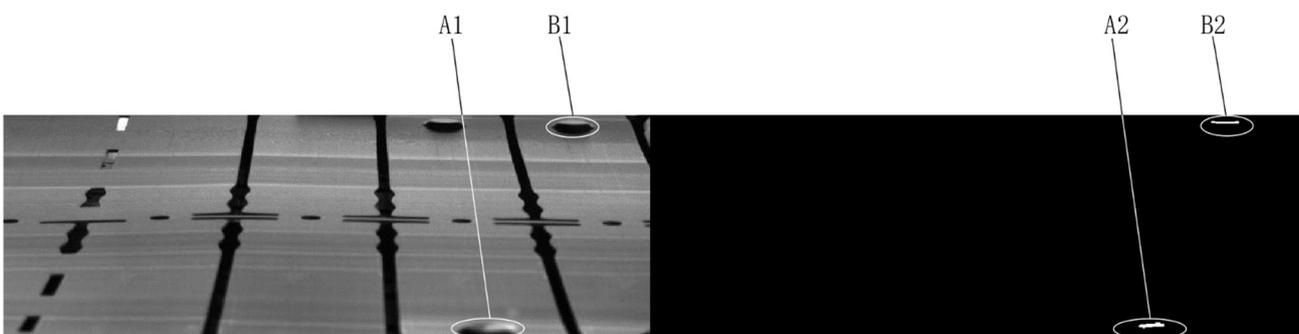
The registered image  $Im_{r_{gs}}$  is obtained by transforming the image  $Im_{wt}$  into the standard template image  $Im_s$ . Then, the difference image  $Im_{df}$  can be calculated using the following formula:

$$Im_{df}(x, y) = \begin{cases} 1, & |Im_{r_{gs}}(x, y) - Im_s(x, y)| \geq Th_{gr} \\ 0, & |Im_{r_{gs}}(x, y) - Im_s(x, y)| < Th_{gr} \end{cases} \quad (5)$$

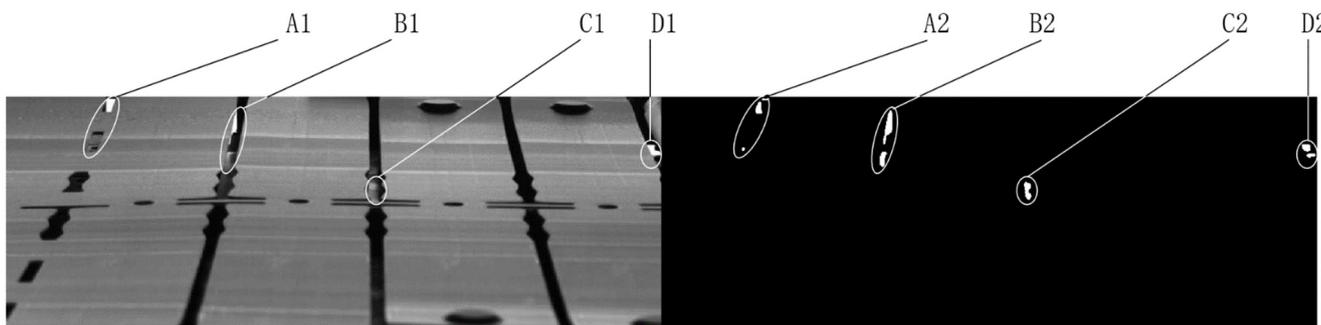
where  $Th_{gr}$  is the Otsu threshold. To enhance the robustness to high-frequency detail interference, a combination of spatial and morphological filtering is performed. (1) Prior to the threshold step, spatial filtering is conducted on  $|Im_{r_{gs}} - Im_s|$  with a Gaussian filter. The purpose is to filter out the high-frequency detail interferences contained in  $Im_{df}$ . (2) Once the difference image is obtained, we filter out the 8 small connected regions in  $Im_{df}$  as they can be considered as interferences that step (1) failed to remove.

The obtained difference image may contain connected regions representing anomalies and/or interferences. All possible types of interference regions and the cause of their presence in the difference image  $Im_{df}$  are discussed in detail:

- (a) As the actual production environment is versatile and complicated, the collected images often contain local bright spots owing to the partial reflection of the workpiece (see region A1 in Figure 2). These local bright spots also cause the generation of interference regions in the difference image (see region A2 in Figure 2).
- (b) Local elastic deformation of the workpiece owing to external forces during production and inherent errors in image registration methods may lead to a complex local transformation between certain regions in  $Im_{r_{gs}}$  and their corresponding regions in  $Im_s$ , which may result in interference regions, such as region B2 in Figure 2. The transformation model between the region pairs is simplified as a translation model when the high-order distortion terms can be omitted.
- (c) The background exposed through the hole structure on the workpiece may also result in interference regions in the difference image (see Figure 3). A background image of the workpiece is captured to remove such an interference region.



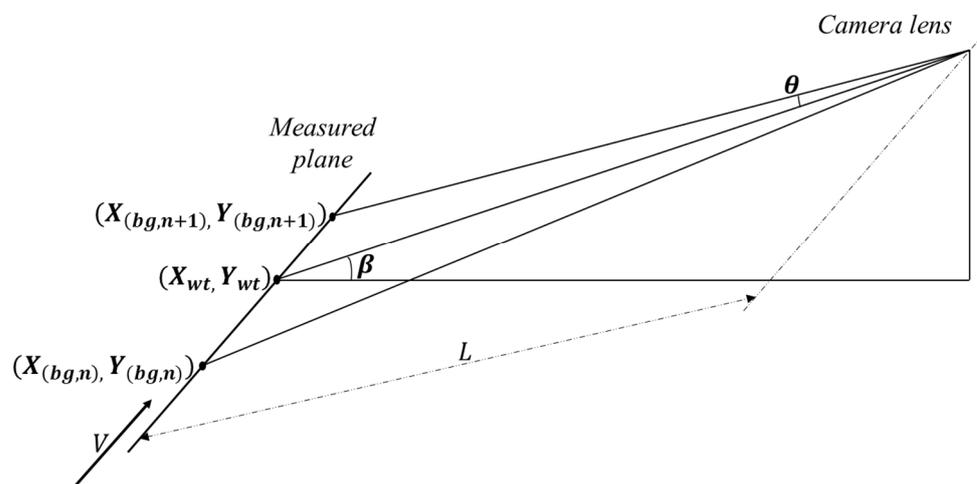
**Figure 2.** Example of a local bright spot region (A1), a bright-spot-caused interference region (A2), a local elastic deformation region (B1), and an elastic-deformation-caused interference region (B2).



**Figure 3.** The hole structures in the workpiece (A1–D1) and corresponding interferences in the difference image (A2–D2).

Due to the constraints of the real scene in the production of stamping progressive dies, it is difficult for the camera to shoot from directly above the stamping parts to achieve the highest resolution and reduce image distortion, thereby achieving the minimum information loss according to the Shannon sampling theorem. In practice, shooting can only occur from the side of the stamping parts. The geometric modeling [52] of the experimental scene is shown in Figure 4. It is assumed that the target’s movement speed is  $V$ , and the direction of speed is shown in the figure. The camera shoots at a downward angle of  $\beta$  degrees, and  $L$  is the distance from the lens to the target point. Therefore, in this model, the theoretical maximum viewing angle change  $\theta$  in the image to be measured and the corresponding background library is calculated as shown in Formula (6).

$$\theta = \arctan \frac{V \cdot T}{2L} \tag{6}$$



**Figure 4.** A sketch map of the theoretical maximum angle of view change,  $\theta$ .

When the height of an object in the scene is  $h$ , then the maximum occlusion pixel  $pt$  caused by the perspective change due to side shooting along the direction of motion is calculated as shown in Formula 7.

$$pt = \frac{h \cdot \tan \theta}{\sin \beta} \cdot r = \frac{hVT}{2L \sin \beta} \cdot r \tag{7}$$

In this model,  $pt$  represents the maximum error, and  $r$  denotes the image resolution. Given that  $\beta$  is small,  $\sin \beta$  can be approximated as 1, and since  $L$  is significantly larger than  $VT$ , the value of  $pt$  tends to be small. This implies that the error introduced by  $pt$  is within

a tolerable range and can be further minimized through subsequent morpho-logical steps. Consequently, the equivalent diameter (in pixels) for anomaly detection is as follows:

$$d = \sqrt{(pt + e_{vib} + e_{reb})^2 \cdot \sin\beta} \quad (8)$$

In this context,  $e_{vib}$  represents the error attributed to vibration, while  $e_{reb}$  signifies the error resulting from registration. Utilizing this method can substantially streamline the scene and address the issue at hand.

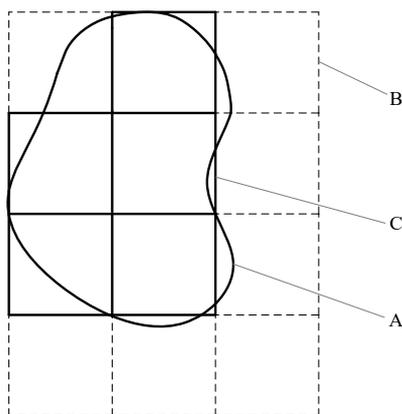
## 2.2. SSAD Construction

As previously discussed, there are often interference regions in  $Im_{df}$ , necessitating further investigation to ascertain whether the connected regions in the difference image represent actual anomalies. Based on the fact that there are two corresponding regions in  $Im_{rgs}$  and  $Im_s$  (candidate region and reference region) for a connected region in  $Im_{df}$ , we compare the descriptors describing the candidate regions with those describing the reference regions to complete this task. However, the existing feature descriptors are for image features, and they are of fixed shapes and sizes. If these descriptors are utilized directly to describe the characteristics of anomalies, information other than that of anomalies will be included. If the scale of the anomaly is small, its information that is reflected in the descriptor will be reduced relative to the total information contained in the descriptor. Therefore, descriptors of fixed sizes and shapes are less distinctive when describing the characteristics of anomaly regions. To overcome this issue, we propose a distinctive SSAD for the connected region. By calculating the matching distances between the SSADs of candidate and reference re-gions, we can identify connected regions that contain anomalies.

To construct the SSAD, we identified the left vertical tangent and top horizontal tangent of a connected region. A rectangle was then formed with the intersection point of the two tangents as a vertex. This rectangle, which could contain the connected region and sides in the vertical or horizontal direction, had a size that was an integer multiple of  $3s$ . Here,  $s$  was the sampling step, which depended on the area of the connected region  $A$  and was determined as follows

$$s = \left\lceil \frac{A}{4A_0} \right\rceil + 1 \quad (9)$$

We propose a method where the rounding function is denoted as  $\lceil \cdot \rceil$ , and we suggest a constant  $A_0$ , typically set to 30 in this paper. We further partitioned the rectangle into square sub-regions of  $3s$  size in a regular manner. Sub-regions whose centers did not fall within the connected region were eliminated, as depicted in Figure 5. The sample points from the connected region were derived by identifying a grid of  $3 \times 3$  sample points in each remaining sub-region. Given the existence of a corresponding connected region in  $Im_{df}$  for both the candidate and reference regions, from which the SSADs were to be extracted, and the definite positional relationship among these connected regions, we computed the sample points for each connected region in  $Im_{df}$ . This allowed us to obtain the sample points of their corresponding connected regions.



**Figure 5.** Sketched map depicting splitting of connected region into  $3 \times 3$  square sub-regions: (A) connected region, (B) rectangle, and (C) remaining sub-region.

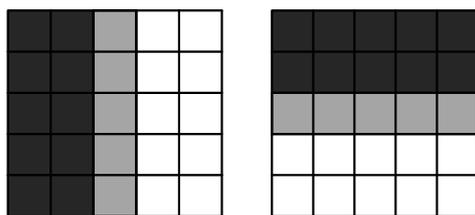
The above-mentioned procedures allocated sample points to the connected region. The next phase involved applying a Gaussian smoothing filter ( $\sigma = 0.01 \times A$ ) to the connected region, followed by determining the response at each sample point using the Haar wavelet filter, as shown in Figure 6. We denoted the operator response in the x-direction as  $dx$  and that in the y-direction as  $dy$ . Subsequently, we aggregated the responses  $dx$  and  $dy$  and their absolute values across each sub-region to yield  $\sum dx, \sum |dx|, \sum dy, \sum |dy|$ , as depicted in Figure 7. For each sub-region,  $\sum dx, \sum |dx|, \sum dy, \sum |dy|$  formed a 4D vector  $\alpha$

$$\alpha = (\sum dx, \sum |dx|, \sum dy, \sum |dy|)^T. \tag{10}$$

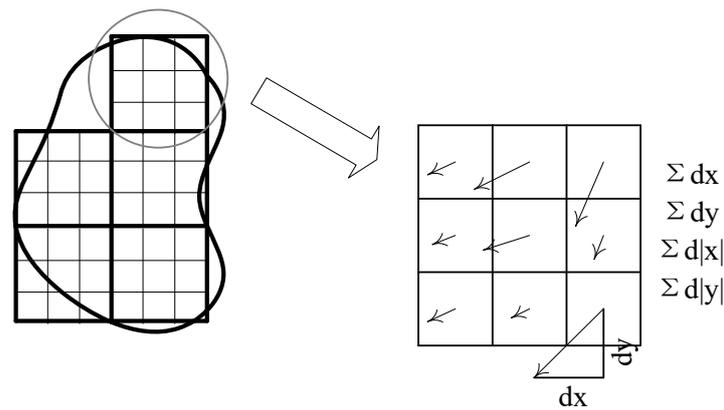
To counteract the influence of local bright spots, the 4D vector  $v$  was converted into a unit vector  $e$ :

$$e = \frac{\alpha}{\|\alpha\|} \tag{11}$$

The SSAD was computed by amalgamating the 4D vectors from all sub-regions of a connected region into an extended vector,  $\beta$ . If a connected region comprised  $n$  sub-regions, its SSAD, which was composed of the 4D vector extracted from each of its sub-regions, was a  $4nD$  vector.



**Figure 6.** Haar wavelet filters for response  $dx$  (left) and response  $dy$  (right). The dark parts have a weight of  $-1$ , while the gray part is  $0$ , and the light part is  $+1$ .



**Figure 7.** Sketch map of descriptor construction.

There are several reasons for choosing the aforementioned sampling steps for the connected regions of different areas and performing smoothing filtering when constructing the SSAD. Firstly, the cost of calculation of descriptor extraction can be decreased. Secondly, the lower resolution of the connected region of a larger area means it ignores certain noise details, and this increases the stability of the algorithm.

As mentioned previously, the error of the image registration and local elastic deformation may result in a spatial offset between two connected regions in  $Im_{rgs}$  and  $Im_s$  corresponding to a region on the workpiece. The translation of the reference region is undertaken to compensate for such an offset. The key to solving this problem is to identify the translation quantity. SURF uses the method proposed by Brown and Lowe [58] to determine the interpolated locations of the points of interest. Their approach uses the Taylor expansion of the scale-space function,  $D(x, y, \sigma)$ :

$$D(\mathbf{X}) = D + \frac{\partial D^T}{\partial \mathbf{X}} \mathbf{X} + \frac{1}{2} \mathbf{X}^T \frac{\partial^2 D}{\partial \mathbf{X}^2} \mathbf{X}, \quad (12)$$

where the origin of  $D(x, y, \sigma)$  is at the sample point, and  $\mathbf{X} = (x, y, \sigma)$  is the offset from the origin.  $\tilde{\mathbf{X}}$  is obtained by calculating the derivative of  $D$  and setting it to zero, yielding

$$\tilde{\mathbf{X}} = -\frac{\partial^2 D^{-1}}{\partial \mathbf{X}^2} \frac{\partial D}{\partial \mathbf{X}'} \quad (13)$$

where  $\tilde{\mathbf{X}}$  is the offset from the sample point to the point of interest. If all elements of  $\tilde{\mathbf{X}}$  are greater than 0.5, the point of interest is near another sample point. The location of the point of interest is determined by adding the offset  $\tilde{\mathbf{X}}$  and the sample point location. According to Brown's method [58], the theoretic error of the point-of-interest location is less than 0.5 in any dimension. The maximum registration error is obtained by substituting  $(0.5, 0.5, 0)^T$  into Equation (4), yielding

$$\nabla \mathbf{X}_{rgs} = (0.5a_{11} + 0.5a_{12}, 0.5a_{21} + 0.5a_{22})^T. \quad (14)$$

The maximum reference region translation quantity,  $\mathbf{X}_{It}$ , is obtained by substituting  $\nabla \mathbf{X}_{rgs}$  into the following equation:

$$\mathbf{X}_{It} = \frac{1}{\rho} \mathbf{X}_t + \nabla \mathbf{X}_{rgs}, \quad (15)$$

where  $\rho$  is the realistic length represented by a pixel, and  $\mathbf{X}_t$  is the equivalent local translation quantity of the local elastic deformation quantity. The smallest step of the reference region translation is set to one. In particular, the translation quantities of

the translated connected regions with respect to the original reference region can make up the set  $S_T: \{(u, v) | -x_{It} - 1 \leq u \leq x_{It} + 1, -y_{It} - 1 \leq v \leq y_{It} + 1, u, v \in Z\}$ , where  $(x_{It}, y_{It})^T = \mathbf{X}_{It}$ . Following the reference region translation operation for the set of the connected regions, the SSAD of each connected region should be calculated. The key to this problem is to identify the sample points of each connected region. The coordinate of the sample point of the connected region  $X$  is expressed as the sum of the original reference region,  $\mathbf{X}_r$ , and the translation quantity:

$$\mathbf{X} = \begin{pmatrix} u \\ v \end{pmatrix} + \mathbf{X}_r \quad (u, v) \in S_T. \tag{16}$$

For a reference region, a set of SSADs (reference descriptor set) can be obtained after calculating the SSAD of each connected region produced thereby.

To remove the interference regions in  $Im_{df}$  caused by the hole structures, for the connected region in  $Im_{df}$ , we compare the characteristics of the two corresponding regions in  $Im_{rgs}$  and the background image. For this purpose, we calculate the SSAD of the corresponding region in the background image (background descriptor) of the connected region in  $Im_{df}$ . This involves the same steps as the construction of the SSAD of the candidate region and reference region. Note that the connected region in  $Im_{df}$  and its corresponding region in the background image have different areas, shapes, and even positions. The coordinate of the sample point of its corresponding regions,  $(u, v)$ , can be obtained by substituting the coordinate of the connected region point in  $Im_{df}$ ,  $(x, y)$ , into the following formula:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \mathbf{T}^{-1} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}, \tag{17}$$

where  $\mathbf{T}$  is the transformation matrix from  $Im_{wt}$  to  $Im_s$ .

### 2.3. t-Distribution-Function-Based Anomaly Region Determination

The previous operation constructed one candidate descriptor, one reference descriptor set, and one background descriptor for a connected region in  $Im_{df}$ . The next step is to infer the formula for anomaly region determination. The notation  $\beta_i^{can}$  can be utilized to represent the candidate descriptor corresponding to the  $i$ 'th connected region in  $Im_{df}$  and  $\beta_{ij}^{rob}$ , the  $j$ 'th reference descriptor corresponding to the  $i$ 'th connected region in  $Im_{df}$ , or the corresponding background descriptor ( $i = 1, 2, \dots, I$ , and  $j = 1, 2, \dots, J + 1$ ;  $I$  is the number of connected regions in  $Im_{df}$ , and  $J$  is the number of reference descriptors corresponding to a connected region in  $Im_{df}$ ). In particular,  $\beta_{ij_0}^{rob}$  corresponding to index  $j_0$  represents the reference descriptor describing the connected region with a translation quantity of 0.

Since the majority of offsets between two connected regions in  $Im_{rgs}$  and  $Im_s$  corresponding to a region of the workpiece are negligible, let random variable  $Y_{ij_0k}$  be the matching distance between the  $k$ 'th 4D vector of  $\beta_i^{can}$  and its corresponding 4D vector of  $\beta_{ij_0}^{rob}$ . According to the law of large numbers,  $Y_{ij_0k}$  obeys normal distribution:

$$Y_{ij_0k} \sim N(\mu, \sigma^2), \tag{18}$$

where  $\sim$  represents "obey". Let  $\bar{Y} = \frac{1}{n} \sum_{i=1}^I \sum_{k=1}^{K_i} Y_{ij_0k}$ , and  $S = \sqrt{\frac{1}{n-1} \sum_{i=1}^I \sum_{k=1}^{K_i} (Y_{ij_0k} - \bar{Y})^2}$ . Then, Formula (19) is obtained by theoretical derivation:

$$\frac{\bar{Y} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\frac{\bar{Y} - \mu}{\frac{\sigma}{\sqrt{n}}}}{\sqrt{\frac{S^2 \cdot (n-1)}{\sigma^2 \cdot (n-1)}}} \sim t(n-1), \tag{19}$$

where  $t(n - 1)$  means t-distribution with a degree of freedom of  $n - 1$ ; thus,

$$P\left(\left|\frac{\bar{Y} - \mu}{\frac{S}{\sqrt{n}}}\right| < t_{\frac{\alpha}{2}}(n - 1)\right) = 1 - \alpha. \tag{20}$$

By substituting the sample value of  $Y_{ij0k}$ ,  $y_{ij0k}$ , and the sample value of  $\bar{Y}$ ,  $\bar{y}$ , into Equation (20), Equation (21) is obtained:

$$\mu < \frac{1}{\sqrt{n}} \sqrt{\frac{1}{n - 1} \sum_{i=1}^I \sum_{k=1}^{K_i} (y_{ij0k} - \bar{y})^2} \cdot t_{\frac{\alpha}{2}}(n - 1) + \bar{y}, \tag{21}$$

where  $n$  is the total number of 4D vectors in the image to be detected;  $K_i$  is the number of 4D vectors in  $\beta_i^{can}$ ;  $\alpha$  is a confidence coefficient; and  $t_{\frac{\alpha}{2}}(n - 1)$  is the t-distribution upper  $\frac{\alpha}{2}$  fractile.

Use the average matching distance over a connected region in  $Im_{df}$  to approximately replace  $\mu$  in Formula (21), and the formula for anomaly region determination is obtained. Given that the descriptor vector of the connected region is comprised of the 4D vectors of its sub-regions at varying positions, and these sub-regions hold different levels of significance, these 4D vectors are weighted using a Gaussian value ( $\sigma = 1.5$  s) at the center of the connected regions during the computation of the average matching distance:

$$E_{ij} = \frac{1}{h} \sum_{k=1}^{K_i} y_{ijk} \cdot \exp\left[-\frac{1}{2} (\mathbf{X}_{ij} - \mathbf{X}_i^{cent})^T \mathbf{\Lambda}^{-1} (\mathbf{X}_{ij} - \mathbf{X}_i^{cent})\right], \tag{22}$$

where  $\mathbf{X}_{ij}$  is the center coordinate of the  $j$ 'th sub-region of the  $i$ 'th connected region in  $Im_{df}$ ;  $\mathbf{X}_i^{cent}$  is the center coordinate of the  $i$ 'th connected region in  $Im_{df}$ ;  $\mathbf{\Lambda} = \begin{pmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{pmatrix}$  is a diagonal matrix;  $y_{ijk}$  is the sample value of the matching distance of the  $k$ 'th 4D vectors, respectively, in  $\beta_i^{can}$  and  $\beta_{ij}^{rob}$ ;  $E_{ij}$  is the average matching distance over the  $i$ 'th connected region in  $Im_{df}$  that is calculated from  $\beta_i^{can}$  and  $\beta_{ij}^{rob}$ . Lastly  $h$  is a normalization factor. This factor is introduced to counteract the influence of the Gaussian weight and standardize the matching distances of SSADs of varying sizes. The calculation of  $h$  is as follows

$$h = \sum_{k=1}^{K_i} \exp\left[-\frac{1}{2} (\mathbf{X}_{ij} - \mathbf{X}_i^{cent})^T \mathbf{\Lambda}^{-1} (\mathbf{X}_{ij} - \mathbf{X}_i^{cent})\right]. \tag{23}$$

For a connected region in  $Im_{df}$ , we obtain a set of average matching distances,  $E_{i,1}, E_{i,2} \dots E_{i,J}, E_{i,J+1}$ . Let  $E_{ij}^{min} = \min\{E_{i,1}, E_{i,2} \dots E_{i,J}, E_{i,J+1}\}$ . According to theoretical derivation, if the  $i$ 'th connected region in  $Im_{df}$  does not represent an anomaly region,  $E_{ij}^{min}$  should satisfy the following equation:

$$E_{ij}^{min} < \frac{1}{\sqrt{n}} \sqrt{\frac{1}{n - 1} \sum_{i=1}^I \sum_{k=1}^{K_i} (y_{ij0k} - \bar{y})^2} \cdot t_{\frac{\alpha}{2}}(n - 1) + \bar{y} \tag{24}$$

Equation (24) is a theoretical formula. In order to increase its robustness and filter out interference, it should be further optimized. For that reason, another important parameter, known as the translation change rate (TCR), is introduced. The TCR  $cr_i$  is obtained using the following formula:

$$cr_i = \frac{E_{i,j_1} + E_{i,j_2}}{\frac{2}{J+1} \sum_{j=1}^{J+1} E_{ij}} \tag{25}$$

where  $E_{i,j_1}$  and  $E_{i,j_2}$  are the maximum average matching distance and the second maximum average matching distance, respectively. The matching distance of the actual anomaly is almost constant when the translation quantity changes. Consequently, an actual anomaly has a small  $cr$  value. The matching distance of the normal region has the minimum value

on the optimal translation quantity and increases sharply when the translation quantity deviates from the optimal translation quantity; thus, the normal region has a large  $cr$  value. This characteristic can be utilized to optimize Formula (24) by adding self-suppression.

In Formula (24), a few  $y_{ij0k}$  values come from the anomaly region. In order to reduce their impact on the accuracy of Formula (24),  $y_{ij0k}$  values from normal regions should be given a larger weight than those from anomaly regions. For this purpose, Formula (24) can be inferred:

$$\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n-1} \sum_{i=1}^I \sum_{k=1}^{K_i} (y_{ij0k} - \bar{y})^2} \cdot t_{\frac{\alpha}{2}}(n-1) + \bar{y} = \sqrt{\frac{1}{n-1} (\bar{y}^2 - \bar{y}^2)} \cdot t_{\frac{\alpha}{2}}(n-1) + \bar{y} \quad (26)$$

where  $\bar{y}^2 = \frac{1}{n} \sum_{i=1}^I \sum_{k=1}^{K_i} Y_{ij0k}^2$ . At the same time, two weight coefficients are defined:

$$Weight_i^1 = \frac{cr_i}{\sum_{i=1}^I cr_i} \quad (27)$$

and

$$Weight_i^2 = \frac{(cr_i)^2}{\sum_{i=1}^I (cr_i)^2} \quad (28)$$

$\bar{y}$  and  $\bar{y}^2$  in Equation (26) are optimized as

$$\bar{z} = \frac{1}{n} \sum_{i=1}^I \sum_{k=1}^{K_i} Y_{ij0k} Weight_i^1 \quad (29)$$

and

$$\bar{z}^2 = \frac{1}{n} \sum_{i=1}^I \sum_{k=1}^{K_i} Y_{ij0k}^2 Weight_i^2 \quad (30)$$

respectively. By using t-distribution and self-suppression optimization, the interference regions' impact is eliminated while determining the actual anomaly, which doubly enhances the robustness and accuracy of the method for anomaly detection.

Finally, through combining Equations (24), (26), (29), and (30), the formula for anomaly determination is obtained:

$$E_{ij}^{min} < \sqrt{\frac{1}{n-1} (\bar{z}^2 - \bar{z}^2)} \cdot t_{\frac{\alpha}{2}}(n-1) + \bar{z}, \quad (31)$$

That is, if a connected region in  $Im_{df}$  does not satisfy Formula (31), it is an anomaly region.

### 3. Comparative Experiments

In this section, we validate the advantages of our proposed method in the monitoring of progressive die stamping production through comparative experiments. Firstly, the experimental implementation is introduced in Section 3.1. In Section 3.2, we compare our proposed descriptor with a widely used descriptor, demonstrating the advantages of our method among non-learning methods. In Section 3.3, we compare our method with several popular deep-learning-based anomaly detection methods. The experimental results show that our method has competitive advantages compared to deep-learning-based methods.

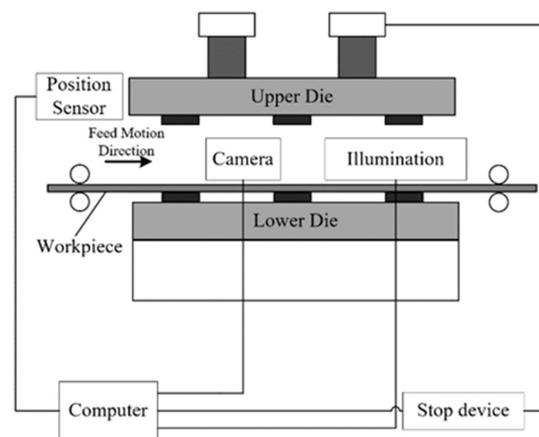
#### 3.1. Implementation of Proposed Method

The 800 T multi-station progressive die, a sophisticated and productive stamping die capable of executing stamping, bending, drawing, forming, and turning within a single die set, was utilized as the test subject to validate the aforementioned algorithm. This die can effectively produce a variety of complex parts. The ROI of the 800 T multistation progressive die production line can be considered a periodic motion scene. The manufacturing process is often plagued by anomalies such as foreign body splashes (processed scraps, spitballs, stains) and machine part loosening, leading to damaged workpieces, waste products, and even equipment damage and malfunction, resulting in substantial economic losses. As a

result, monitoring for anomalies in processing equipment has emerged as a crucial strategy for maintaining normal operations.

The hardware of the detection system consists of a CMOS camera, a computer (Windows system), a planar light source, an auxiliary control system, and corresponding support equipment (see Figure 8). The core program of the detection system runs on the MATLAB 9.6.0 system. The camera is capable of producing 1.3 megapixel 60 fps grayscale images. The camera lens has a focal length of 25 mm, and the working distance  $L$  in this experiment was approximately 1.8 m. Four stations were chosen for detection from the collected images, with a detection area of  $851 \times 371$  pixels and a field of view of approximately  $0.65 \times 0.8$  m<sup>2</sup>. Due to the scene's space constraints, the camera's tilt angle  $\beta$  is relatively small, with  $\beta$  approximately equal to  $21^\circ$  in this experiment. In this scenario, the geometric relationship between pixels and reality on the X-axis and Y-axis is shown in Equation (32):

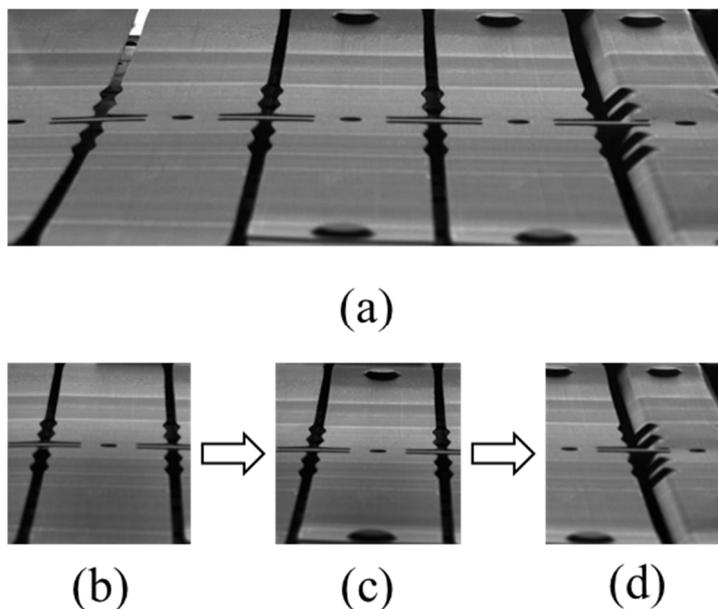
$$\begin{aligned} \text{X-axis: } & 650 \text{ mm}/851 \text{ pixel} \approx 0.76 \text{ mm/pixel} \\ \text{Y-axis: } & \sin(21^\circ) \times 800 \text{ mm}/371 \text{ pixel} \approx 0.77 \text{ mm/pixel} \end{aligned} \quad (32)$$



**Figure 8.** Configuration of detection system on 800 T multi-station progressive die.

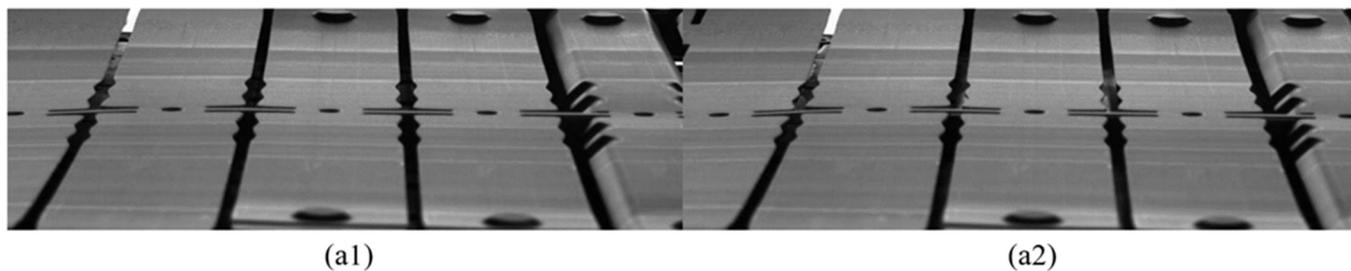
At this focal length, the corresponding relationship between the pixels of the focal image and the actual geometric size is  $\sim 0.8$  mm per pixel. At this specific focal length, the pixel-to-actual geometric size correspondence in the focal image is approximately 0.8 mm per pixel. The camera is capable of capturing 30 images per second, with an interval of roughly 35 ms between each image and an exposure time of 3 ms. The detection environment, being a closed space, exhibits high resilience to changes in illumination.

The sheet metal workpiece to be processed is shown in Figure 9. Figure 9a displays a part of the entire sheet metal processing process, where three states appear simultaneously in Figure 9a: Figure 9b is the workpiece to be punched, Figure 9c is the state after punching, and Figure 9d is the bending state. The sheet metal workpiece moves to the right as a whole during processing, and the overall processing process is cyclical.

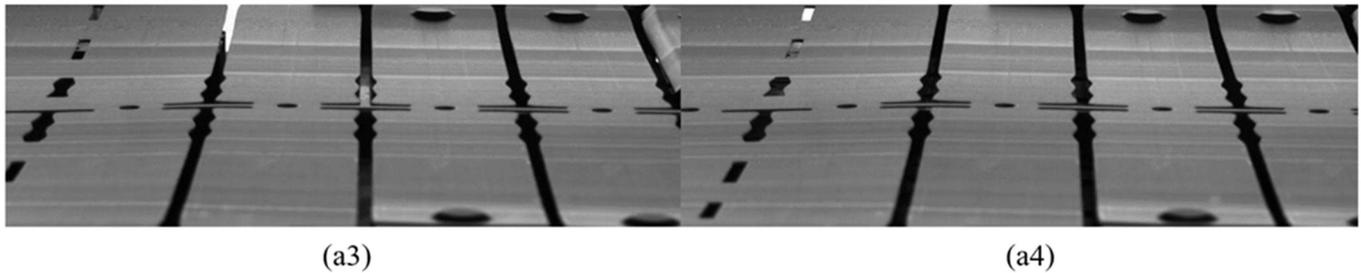


**Figure 9.** Progressive die stamping production process. (a) displays a part of the entire sheet metal processing process (b) is the workpiece to be punched, (c) is the state after punching, and (d) is the bending state.

As discussed in Section 2, the construction of the standard template image library should precede the detection work. Image acquisition and detection are only possible when the die is open, limiting the detection to the time frame when a detectable image can be captured. The multi-station progressive die has an operating cycle of 3.5 s, and the effective image acquisition time is 1.4 s. Due to the fact that the detected area is completely stationary at the beginning and end of this cycle, and the camera exposure time is 3 ms, blur caused by vibration during shooting can be omitted. During this time, 30 images are captured with equal time intervals, then pre-processed, and finally saved as a standard template image library. Figure 10 illustrates part of the standard template image library.



**Figure 10.** *Cont.*



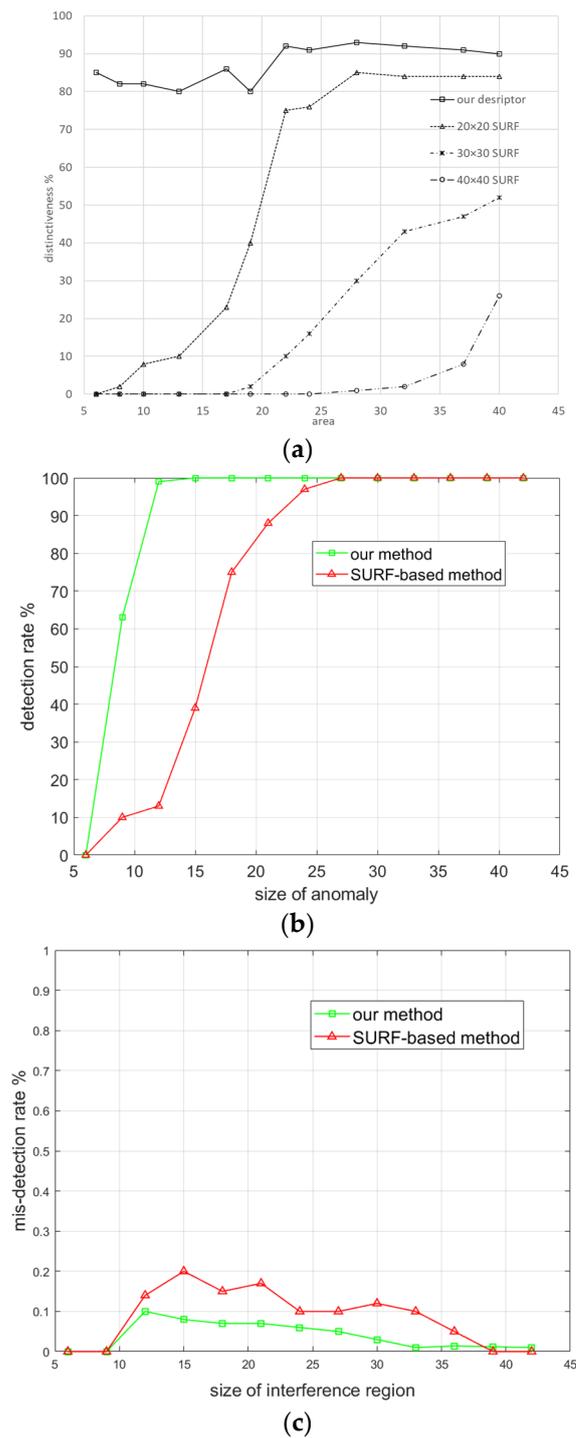
**Figure 10.** Images in standard template image library with corresponding numbers (a1) 7, (a2) 14, (a3) 21, and (a4) 28.

### 3.2. Comparative Experiments with SURF

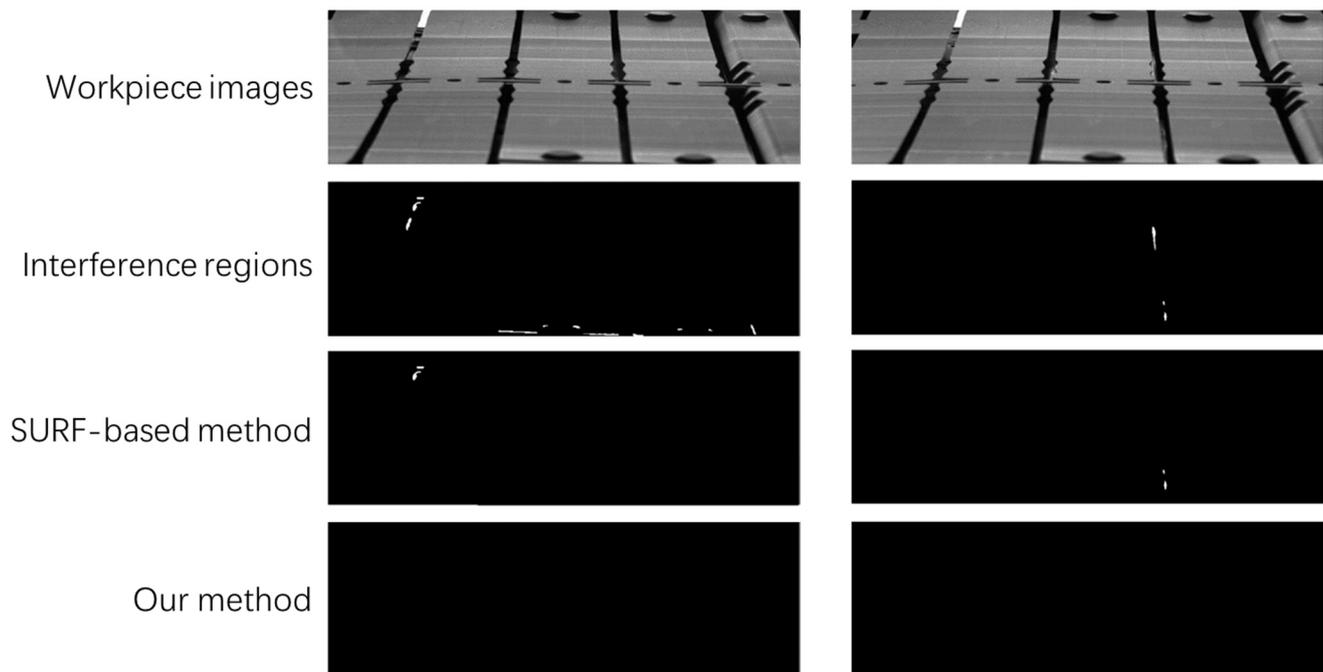
The SURF algorithm, short for Speeded-Up Robust Features, is an improvement over the SIFT operator. While it retains the excellent performance characteristics of SIFT, it addresses the high computational complexity and long computation time associated with SIFT. The SURF algorithm enhances the extraction of interest points and their feature vector descriptions, thereby speeding up computation. The specific steps of the SURF algorithm include constructing the Hessian matrix and calculating the eigenvalue  $\alpha$ , building a Gaussian pyramid, locating feature points, determining the main direction of feature points, and constructing the feature descriptor. These steps ensure that SURF maintains the robustness of SIFT while significantly improving computational efficiency.

In the experiments, our descriptor is compared to the SURF descriptors of different sizes, and performance evaluation is carried out on the die image set. Due to the lack of abnormal samples, we manually added external objects of different sizes to the progressive die and then took photos of the stamped parts with the foreign objects on them, to evaluate the distinctiveness of these descriptors. Anomalous sources may introduce foreign objects through various means, including but not limited to, the dispersion of extraneous materials, the spillage of waste, the de-tachment of components, the inclusion of pollutants in raw materials, and significant plastic deformation of the raw materials. Throughout the stamping process, the components consistently exhibited minor deformations. The distinctiveness scores of the different descriptors are shown in Figure 11a. This score is the ratio of the average matching distance of the anomaly region to ten times the average matching distance of the normal region. We also carried out experiments for detection rate and misdetection rate evaluation on the die image set, comparing our method to the SURF-based method presented in [51]. The results are illustrated in Figure 11b,c.

Our method clearly outperforms the SURF-based method in detecting anomalies when the size of anomalies is less than 20 pixels. These very good performances can be explained by the fact that the SSAD contains less additional information and has a higher distinctiveness score. The SURF-based method has a relatively high misdetection rate. This can be traced back to its subpar performance in eliminating hole interferences, as illustrated in Figure 12. The figure presents four rows: the first row displays the workpiece under inspection, the second row shows the differential image with the interference area, the third row depicts the differential image post interference area filtering using the SURF-based method, and the fourth row presents the image after interference area filtering using our method. It is evident that the SURF-based method struggles with filtering out interference outside the hole.



**Figure 11.** Comparison of methods applied on the die image set. Distinctiveness scores when increasing the sizes of foreign objects (a). Our descriptor's score is greater than 80% for different sizes. Detection rate and mis-detection rate for foreign objects of different sizes (b,c). Our method outperforms the SURF-based method when the foreign objects are relatively small.



**Figure 12.** Comparison of our method with SURF.

The error check rate arises from the theoretical error introduced using the rigid body translation model to approximate the local elastic deformation and by using the affine transformation model to approximate the perspective transformation model. Furthermore, the discrete integer translation step is utilized to compensate for the registration error and elastic deformation, which introduces a discretization error. This type of error is a different source of the error check rate. The detection accuracy of this algorithm is affected by factors such as image resolution, camera tilt angle, and focal length. The resolution of the image is fixed and depends on the camera and lens selection. Because of factors such as registration error (about 1 pixel), machine vibration error (up to about 3 to 5 pixels), and the inherent error of the method (less than 1 pixel according to Equation (7)), according to Equation (8), the equivalent diameters detected under these conditions are

$$\sqrt{(1 + 1 + 4)^2 \cdot \sin 21^\circ} \approx 3.6 \text{ pixel} \approx 3 \text{ mm} \quad (33)$$

As mentioned previously, five images were collected for testing in one cycle, and the detection of these five images should be completed before the next cycle to avoid production accidents caused by anomalies, whereby the effective detection time was 1.4 s. Therefore, the detection time of each image should be controlled within 280 ms. The algorithm had a detection time of 32 ms for an image without any optimization. The computer had an Intel i7-10700 CPU processor (Intel, Oregon, USA) with a clock speed of 2.9 GHz, 32 GB of RAM, and a 64-bit operating system. The core program ran under MATLAB 9.6.0, and the image resolution was  $851 \times 371$  pixels; therefore, real-time detection was realized.

### 3.3. Comparison with the Methods Based on Deep Learning

We established a dataset that includes 500 images without anomalies and 10 images with anomalies, where the pixel positions of the anomalies have been manually annotated. We applied random rotation, cropping, and flipping to augment the data. After data augmentation, the number of images without anomalies reached 4000. The dataset also includes 500 images with anomalies and manually annotated pixel-level labels for training.

For training, we employed a batch size of 32 for training and an Adam optimizer with an initial learning rate of 0.0005. The cosine annealing strategy was utilized to adjust the learning rate, with the penalty parameter  $\beta$  set to 0.1.

The proposed approach was implemented using Pytorch 1.7.0 and executed on a computer equipped with 32 GB of RAM, an Intel i7-10700 2.9 GHz CPU, an NVIDIA RTX 3080 GPU, and an Ubuntu 20.04 operating system.

We trained four standard segmentation and defect detection methods, namely SegNet [59], U-Net++ [60], MobileNetV2+DeepLabV3 [61], and PGA Net [62]. After the training was completed, the four learning-based methods were compared with our method on the same dataset. The test dataset consists of a total of 2000 images, of which 1987 are without anomalies, and 13 are anomalous. We used accuracy, miss rate, and time consumption to measure various methods. Accuracy refers to the proportion of correctly identified images to the total number of images, miss rate refers to images that were judged as normal but were actually abnormal, and time consumption refers to the time required to process a single image. The comparison results are shown in Table 1.

**Table 1.** Comparison of test results using different methods.

Method	Number of Images Judged as Normal		Number of Images Judged as Abnormal		Accuracy (%)	Miss Rate (%)	Time Consumption (%)
	TRUE	FALSE	TRUE	FALSE			
Ours	13	3	1984	0	99.85	0.00	32
SegNet	10	4	1983	3	99.65	0.15	57
U-Net++	13	5	1982	0	99.75	0.00	37
MobileNetV2+DeepLabV3	11	4	1985	0	99.8	0.00	33
PGA Net	11	3	1980	6	99.55	0.30	52

From the results, it can be seen that our method has achieved a high detection accuracy rate compared to several deep-learning-based methods, and our method has a miss rate of 0. In terms of time, all methods can achieve real-time monitoring in the stamping progressive die production process. However, the difference is that our method only needs 30 images in the image library to achieve extremely high accuracy monitoring.

Figure 13 a visualization of the comparison of various methods. Our method identifies anomalies through differential filtering with the proposed descriptor, while other methods do so through semantic segmentation. The first, second, and third columns display three types of anomalies, i.e., different foreign objects appearing on the workstation. The fourth column displays normal images.

Therefore, the monitoring method we proposed exhibits superior performance in scenarios with strong periodic characteristics such as stamping progressive die production, compared to deep-feature methods based on neural networks. This is because it requires a small image library that is easy to prepare and does not need to undergo the data processing and training processes required by deep learning paradigms. Moreover, it meets the demands of production in terms of accuracy and real-time performance.



**Figure 13.** Visualization of the comparison of various methods.

#### 4. Conclusions

This paper proposes a method for detecting anomalies in periodic motion scenes, which can be widely applied to production lines with these types of scenes. The proposed method has the following characteristics:

- (1) The proposed SSAD for region description breaks the inherent mode of the traditional descriptor. Its adaptability to the shape and size of the anomaly region makes sure it is more distinctive. In constructing the SSAD, adaptive resolutions are used to describe the anomaly region, which reduces the computational cost of the feature extraction calculation, ignores high-frequency noise interferences, and improves the signal-to-noise ratio of the descriptor.
- (2) This study introduced a novel method based on t-distribution for anomaly detection, which abandoned the traditional empirical theoretical threshold, showing a higher robustness. Meanwhile, self-suppression optimization based on TCR was used in this study, which drastically reduced the misdetection rate.
- (3) The maximum translation quantity was inferred to filter out local elastic deformation, and the background descriptor was constructed to eradicate the impacts of backgrounds exposed through holes, reducing the misdetection miss rate to 0.0%.
- (4) The proposed method outperforms the deep-feature method as it necessitates only a minimal number of images to construct an image library, and the level of detection can achieve results comparable to those of prevalent neural networks. At the same time, our method does not require paradigms such as knowledge transfer, pre-training, and fine-tuning of neural networks, making the preprocessing process simpler.

In this study, we used a progressive die to test this method. The experimental results show that the proposed algorithm can achieve comparable or even superior performance in terms of the anomaly detection rate compared to its counterparts and is superior in terms of the misdetection rate. Generally, processing equipment with periodic motion scenes, such as dynamic injection molds and printing machines, can be monitored using this method: first, a standard template image library is constructed during the periodic motion process

of the mold, then the descriptor is constructed, and finally an anomaly region is determined based on the T-distribution function.

The focus of future work should be on conducting online monitoring tests of the method we proposed in the production of various processing equipment with periodic motion scenes. This will be the work of the next stage.

**Author Contributions:** F.M. was in charge of the whole thesis and revised the final manuscript; L.M. wrote the initial manuscript. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by the Key R&D Projects of Hebei Province (No. 213110301D).

**Institutional Review Board Statement:** The study did not require ethical approval.

**Informed Consent Statement:** This study did not involve humans.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare that they have no competing interests.

## References

- Xia, F.Z.; Sun, H.B.; Wei, H.G. Rolling Textures in BCC Metals: A Biaxial Stress Texture Theory and Experiments. *Phys. Met. Met.* **2021**, *122*, 710–717. [\[CrossRef\]](#)
- Wan, M.M.; Li, F.G.; Yao, K.; Song, G.; Fan, X.G. Theory, Method and Practice of Metal Deformation Instability: A Review. *Materials* **2023**, *16*, 2667. [\[CrossRef\]](#) [\[PubMed\]](#)
- Yang, Y.; Hinduja, S. Sequence planning of sheet metal parts manufactured using progressive dies. *Int. J. Adv. Manuf. Technol.* **2022**, *124*, 2199–2214. [\[CrossRef\]](#)
- Zhou, C.; Liu, K.; Zhang, X.; Zhang, W.; Shi, J. An automatic process monitoring method using recurrence plot in progressive stamping processes. *IEEE Trans. Autom. Sci. Eng.* **2015**, *13*, 1102–1111. [\[CrossRef\]](#)
- Leptidis, S.; Papageorgiou, D.G.; Medrea, C.; Chicinas, I. Failure analysis of an EDM machined mold-printing die used for the production of truck spare parts. *Eng. Fail. Anal.* **2016**, *61*, 62–68. [\[CrossRef\]](#)
- Quadir, A.; Jha, K.; Haneef, S.; Sivaraman, A.K.; Tee, K.F. A review on data-driven quality prediction in the production process with machine learning for industry 4.0. *Processes* **2022**, *10*, 1966.
- Sah, S.; Mahayotsanun, N.; Peshkin, M.; Cao, J.; Gao, R.X. Pressure and draw-in maps for stamping process monitoring. *J. Manuf. Sci. Eng.* **2016**, *138*, 91005. [\[CrossRef\]](#)
- Mahayotsanun, N.; Sah, S.; Cao, J.; Peshkin, M.; Gao, R.X.; Wang, C.T. Tooling-integrated sensing systems for stamping process monitoring. *Int. J. Mach. Tools Manuf.* **2009**, *49*, 634–644. [\[CrossRef\]](#)
- Xu, Y.; Ge, M. Hidden Markov model-based process monitoring system. *J. Intell. Manuf.* **2004**, *15*, 337–350. [\[CrossRef\]](#)
- Li, Z.; Liu, R.; Wu, D. Data-driven smart manufacturing: Tool wear monitoring with audio signals and machine learning. *J. Manuf. Process.* **2019**, *48*, 66–76. [\[CrossRef\]](#)
- Kim, J.; Huang, Q.; Shi, J.; Chang, T.S. Online Multichannel Forging Tonnage Monitoring and Fault Pattern Discrimination Using Principal Curve. *J. Manuf. Sci. Eng.* **2006**, *128*, 944–950. [\[CrossRef\]](#)
- Jing, L.; Tingting, D.; Dan, S.; Chunbo, X. A review on surface defect detection. *J. Front. Comput. Sci. Technol.* **2014**, *8*, 1041–1048.
- Ghorai, S.; Mukherjee, A.; Gangadaran, M.; Dutta, P.K. Automatic defect detection on hot-rolled flat steel products. *IEEE Trans. Instrum. Meas.* **2013**, *62*, 612–621. [\[CrossRef\]](#)
- Xie, L.; Huang, R.; Gu, N.; Cao, Z. A novel defect detection and identification method in optical inspection. *Neural Comput. Appl.* **2014**, *24*, 1953–1962. [\[CrossRef\]](#)
- Liu, K.; Wang, H.; Chen, H.; Qu, E.; Tian, Y.; Sun, H. Steel surface defect detection using a new haar-Weibull-variance model in unsupervised manner. *IEEE Trans. Instrum. Meas.* **2017**, *66*, 2585–2596. [\[CrossRef\]](#)
- Truong, M.T.N.; Kim, S. Automatic image thresholding using Otsu's method and entropy weighting scheme for surface defect detection. *Soft Comput.* **2018**, *22*, 4197–4203. [\[CrossRef\]](#)
- Martinez, P.; Ahmad, R.; Al-Husseini, M. A vision-based system for pre-inspection of steel frame manufacturing. *Autom. Constr.* **2019**, *97*, 151–163. [\[CrossRef\]](#)
- Lin, T.K. An Adaptive Vision-Based Method for Automated Inspection in Manufacturing. *Adv. Mech. Eng.* **2014**, *2014*, 616341. [\[CrossRef\]](#)
- Liu, J.P.; Tang, Z.H.; Zhang, J.; Chen, Q.; Xu, P.; Liu, W. Visual perception-based Statistical Modeling of Complex Grain Image for Product Quality Monitoring and Supervision on Assembly Production Line. *PLoS ONE* **2016**, *11*, e0146484. [\[CrossRef\]](#)
- Gamage, P.; Xie, S.Q. A real-time vision system for defect inspection in cast extrusion manufacturing process. *Int. J. Adv. Manuf. Technol.* **2009**, *40*, 144–156. [\[CrossRef\]](#)
- Chen, Q.L.; Wang, D.X.; Lin, J.G. Surface defect detection on stamping and grinding flat parts. *Forg. Stamp. Technol.* **2020**, *45*, 168–174.

22. Leo, M.; Del Coco, M.; Carcagnì, P.; Spagnolo, P.; Mazzeo, P.L.; Distante, C.; Zecca, R. Automatic visual monitoring of welding procedure in stainless steel kegs. *Opt. Lasers Eng.* **2018**, *104*, 220–231. [[CrossRef](#)]
23. Cao, B.F.; Li, J.Q.; Liu, C.F.; Qin, L. Defect detection of nickel plated punched steel strip based on improved least square method. *Optik* **2020**, *206*, 164331. [[CrossRef](#)]
24. Li, B.; Wu, Y.; Guo, F.X.; Qi, J. Real-Time Detection Method for Surface Defects of Stamping Parts Based on Template Matching. *IOP Conf. Ser. Earth Environ. Sci.* **2019**, *252*, 022076. [[CrossRef](#)]
25. Luo, Q.; Fang, X.; Liu, L.; Yang, C.; Sun, Y. Automated Visual Defect Detection for Flat Steel Surface: A Survey. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 626–644. [[CrossRef](#)]
26. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the 3rd International Conference on Learning Representations, San Diego, CA, USA, 7–9 May 2015.
27. Szegedy, C.; Liu, W.; Jia, Y.Q.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A.; Liu, W.; et al. Going deeper with convolutions. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 7–12 June 2015; pp. 1–9. [[CrossRef](#)]
28. He, K.M.; Zhang, X.Y.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [[CrossRef](#)]
29. Shang, L.D.; Yang, Q.S.; Wang, J.N.; Li, S.B.; Lei, W.M. Detection of rail surface defects based on CNN image recognition and classification. In Proceedings of the International Conference on Advanced Communication Technology, ICACT, Chuncheon, Republic of Korea, 1–14 February 2018; pp. 45–51.
30. Sassi, P.; Tripicchio, P.; Avizzano, C.A. A smart monitoring system for automatic welding defect detection. *IEEE Trans. Ind. Electron.* **2019**, *66*, 9641–9650. [[CrossRef](#)]
31. Ma, L.; Xie, W.; Zhang, Y. Blister defect detection based on convolutional neural network for polymer lithium-ion battery. *Appl. Sci.* **2019**, *9*, 1085. [[CrossRef](#)]
32. Akram, M.W.; Li, G.Q.; Jin, Y.; Chen, X.; Zhu, C.G.; Zhao, X.D.; Khaliq, A.; Faheem, M.; Ahmad, A. CNN based automatic detection of photovoltaic cell defects in electroluminescence images. *Energy* **2019**, *189*, 116319. [[CrossRef](#)]
33. Natarajan, V.; Hung, T.Y.; Vaikundam, S.; Chia, L.T. Convolutional networks for voting-based anomaly classification in metal surface inspection. In Proceedings of the IEEE International Conference on Industrial Technology (ICIT), Toronto, ON, Canada, 22–25 March 2017; pp. 986–991. [[CrossRef](#)]
34. Tao, X.; Wang, Z.H.; Zhang, Z.T.; Zhang, D.P.; Xu, D.; Gong, X.Y.; Zhang, L. Wire defect recognition of spring-wire socket using multitask convolutional neural networks. *IEEE Trans. Compon. Packag. Manuf. Technol.* **2018**, *8*, 689–698. [[CrossRef](#)]
35. Yin, X.; Chen, Y.; Bouferguene, A.; Zaman, H.; Al-Hussein, M.; Kurach, L. A deep learning-based framework for an automated defect detection system for sewer pipes. *Autom. Constr.* **2020**, *109*, 102967. [[CrossRef](#)]
36. Feng, C.; Zhang, H.; Wang, H.; Wang, S.; Li, Y. Automatic pixel-level crack detection on dam surface using deep convolutional network. *Sensors* **2020**, *20*, 2069. [[CrossRef](#)] [[PubMed](#)]
37. Xiao, L.; Huang, T.; Wu, B.; Hu, Y.; Zhou, J. Surface defect detection using hierarchical features. In Proceedings of the 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), Vancouver, BC, Canada, 22–26 August 2019; pp. 1592–1596. [[CrossRef](#)]
38. Zipfel, J.; Verworner, F.; Fischer, M.; Wieland, U.; Kraus, M.; Zschech, P. Anomaly detection for industrial quality assurance: A comparative evaluation of unsupervised deep learning models. *Comput. Ind. Eng.* **2023**, *177*, 109045. [[CrossRef](#)]
39. DeTone, D.; Malisiewicz, T.; Rabinovich, A. SuperPoint: Self-supervised interest point detection and description. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 224–236.
40. Araslanov, N.; Roth, S. Self-supervised Augmentation Consistency for Adapting Semantic Segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 15384–15394.
41. Pautrat, R.; Lin, J.-T.; Larsson, V.; Oswald, M.R.; Pollefeys, M. SOLD<sup>2</sup>: Self-Supervised Occlusion-Aware Line Description and Detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA, 20–25 June 2021; pp. 11368–11378.
42. Xu, R.; Hao, R.; Huang, B. Efficient surface defect detection using self-supervised learning strategy and segmentation network. *Adv. Eng. Inform.* **2022**, *52*, 101566. [[CrossRef](#)]
43. Tsai, D.M.; Jen, P.H. Autoencoder-based anomaly detection for surface defect inspection. *Adv. Eng. Inform.* **2021**, *48*, 101272. [[CrossRef](#)]
44. Chow, J.K.; Su, Z.; Wu, J.; Tan, P.S.; Mao, X.; Wang, Y. Anomaly detection of defects on concrete structures with the convolutional autoencoder. *Adv. Eng. Inform.* **2020**, *45*, 101105. [[CrossRef](#)]
45. Givnan, S.; Chalmers, C.; Fergus, P.; Ortega-Martorell, S.; Whalley, T. Anomaly detection using autoencoder reconstruction upon industrial motors. *Sensors* **2022**, *22*, 3166. [[CrossRef](#)]
46. Mishra, P.; Verk, R.; Fornasier, D.; Piciarelli, C.; Foresti, G.L. VT-ADL: A Vision Transformer Network for Image Anomaly Detection and Localization. In Proceedings of the IEEE/IES International Symposium on Industrial Electronics (ISIE), Kyoto, Japan, 20–23 June 2021; pp. 1–6.

47. Wu, H.; Li, B.; Tian, L.; Sun, Z.; Dong, C.; Liao, W. CoRe: Contrastive and Restorative Self-Supervised Learning for Surface Defect Inspection. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 5020112. [[CrossRef](#)]
48. Guo, Y.; Zhu, W.; Jiao, P.; Chen, J. Foreground detection of group-housed pigs based on the combination of Mixture of Gaussians using prediction mechanism and threshold segmentation. *Biosyst. Eng.* **2014**, *125*, 98–104. [[CrossRef](#)]
49. Mittal, A.; Monnet, A.; Paragios, N. Scene modeling and change detection in dynamic scenes: A subspace approach. *Comput. Vis. Image Underst.* **2009**, *113*, 63–79. [[CrossRef](#)]
50. Chen, X.M.; Liao, J.; Li, B.; Chen, Q.M. Foreground detection based on modified ViBe in dynamic background. *Opt. Precis. Eng.* **2014**, *22*, 2545–2552. [[CrossRef](#)]
51. Wang, L.; Xu, Y.; Meng, F. Detection of foreign bodies in periodic motion scenes using shape- and size-adaptive de-scriptors. In Proceedings of the 2021 IEEE International Conference on Power, Intelligent Computing and Systems (ICPICS), Shenyang, China, 29–31 July 2021. [[CrossRef](#)]
52. Wang, Q.; Meng, F.; Huang, Z.; Li, K. Anomaly detection in periodic motion scenes based on multi-scale feature Gaussian weighting analysis. *Meas. Sci. Technol.* **2019**, *30*, 055602. [[CrossRef](#)]
53. Cutler, R.; Davis, L.S. Robust real-time periodic motion detection, analysis, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *22*, 781–796. [[CrossRef](#)]
54. Zitová, B.; Flusser, J. Image registration methods: A survey. *Image Vis. Comput.* **2003**, *21*, 977–1000. [[CrossRef](#)]
55. Bay, H.; Ess, A.; Tuytelaars, T.; Van Gool, L. Speeded-Up Robust Features (SURF). *Comput. Vis. Image Underst.* **2008**, *110*, 346–359. [[CrossRef](#)]
56. Saleem, S.; Bais, A.; Sablatnig, R. A Performance Evaluation of SIFT and SURF for Multispectral Image Matching. In *International Conference Image Analysis and Recognition*; Springer: Berlin/Heidelberg, Germany, 2012; Volume 7324, pp. 166–173.
57. Qiu, P.; Xing, C. Feature based image registration using non-degenerate pixels. *Signal Process.* **2013**, *93*, 706–720. [[CrossRef](#)]
58. Lowe, D.G. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [[CrossRef](#)]
59. Badrinarayanan, V.; Kendall, A.; Cipolla, R. SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *39*, 2481–2495. [[CrossRef](#)]
60. Zhou, Z.; Siddiquee, M.M.R.; Tajbakhsh, N.; Liang, J. Unet++: A Nested U-Net Architecture for Medical Image Segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*; Springer: Berlin/Heidelberg, Germany, 2018; pp. 3–11.
61. Sandler, M.; Howard, A.; Zhu, M.; Zhmoginov, A.; Chen, L.C. MobileNetV2: Inverted Residuals and Linear Bottlenecks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 4510–4520.
62. Dong, H.; Song, K.; He, Y.; Xu, J.; Yan, Y.; Meng, Q. PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection. *IEEE Trans. Ind. Inform.* **2020**, *16*, 7448–7458. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.