Available online at www.sciencedirect.com

ScienceDirect

Procedia Manufacturing 37 (2019) 394–401

Procedia MANUFACTURING

www.elsevier.com/locate/procedia

9th International Conference on Physical and Numerical Simulation of Materials Processing (ICPNS'2019)

# Research on Approaches for Computer Aided Detection of Casting Defects in X-ray Images with Feature Engineering and Machine Learning

Bo Wu, Jianxin Zhou, Xiaoyuan Ji*, Yajun Yin, Xu Shen

*State Key Laboratory of Materials Processing and Die & Mould Technology, Huazhong University of Science & Technology, Wuhan 430074, China*

**Abstract**

X-ray testing has been adopted as the principal non-destructive testing approach to identify defects within a casting component. However, manual detection for X-ray images carried out by operator or expert always tends to be time-consuming, subjective and error-prone. Intelligent inspection techniques based on computer vision, which have been broadly employed in object recognition with promising results in optical natural images, provides a new idea for computer aided detection of casting defects in X-ray images. In this paper, we compare and evaluate several methods, most of which have not been researched for computer aided detection of casting defects in X-ray images and are based on different feature engineering methods and machine learning models, including local binary patterns-SVM, Gabor-SGD, histogram of oriented gradient-random forest and combination among them, to pursue an approach with better performance on detection of casting defects in X-ray images from the our InteCAST dataset. The experimental results demonstrate that the best performance was acquired by LBP feature and an ensemble learning model, which indicates that the approach proposed provides valuable reference for solving the problems in manual detection.



*Keywords:* Computer aided detection; Casting defects; X-ray images; Feature Engineering; Machine learning

* Corresponding author. Tel.: +86-027-87558147; fax: +86-027-87558147.
*E-mail address:* jixiaoyuan@hust.edu.cn


10.1016/j.promfg.2019.12.065

## 1. Introduction

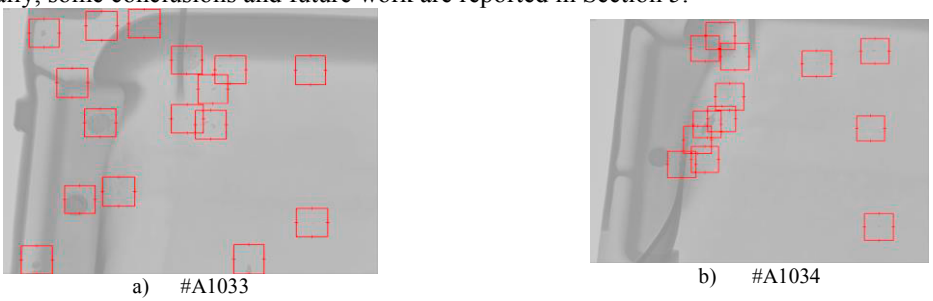
Foundry is the footstone of aerospace, ship, automobile, machinery and other manufacturing industries. Casting defects being harmful to the final product quality during processes always occur [1].Cracks, gas-holes, high-inclusions, low-inclusions, shrinks, shrinkage-holes and incomplete fusion are several kinds of common casting defects.[2]. Failure of critical mechanical components would occur, if those defects are not able to be detected. Several examples of castings with defects are illustrated in Fig. 1. It is a necessary to check every part thoroughly to avoid accidents. Detection defect early is helpful to find defective products as early as possible during manufacturing process and to save time as well as cost [3].

Nondestructive testing technology for castings is a necessary procedure to ensure the quality of products. Among all kinds of non-destructive testing technologies, X-ray testing methods are widely used because they can detect the internal structure of the object to be detected and display in real time. In the process of inspection, at least 10 X-ray images must be collected from each casting to ensure the comprehensive inspection of each part. In many foundry enterprises, manual inspection (one person reviews one person, or even many people review), which is still a critical element in the process, is mostly adopted to carry out quality inspection. The quality of defect detection is related to the operator's technical level, mental state and inspection experience. These uncertainties make the responsible accidents, which could be very expensive for the company, caused by defect omission occur from time to time. In addition, with the increase of castings production, the contradiction between the casting yields and the inefficient manual inspection is particularly prominent and the performance of human-based inspection does not scale well. For these reasons, the application of computer aided X-ray systems naturally emerges as a solution to this problem for the development of foundry enterprises. X-ray systems have not only raised quality through repeated objective inspections and improved processes, but have also increased productivity and consistency by reducing labor costs.

However, computer aided identifying casting defects in X-ray images still remains a challenging task. Cropped images of regions with and without defects are shown in Fig. 2, from some of which it is clear that there are some patterns that can be easily detected (e.g. defects that are irregular in shape and size with bright background and no-defects that are with regular structures). However, the recognition of both classes can be very difficult for low contrast defects because they are very similar to homogeneous no-defects. Taking Fig. 2.a 2,10 (image patch at row 2 column 10 in Fig. 2.a) and Fig. 2.b 2,10 (image patch at row 2 column 10 in Fig. 2. b) as example, the flaw is just a darker spot and the flaws signal is only slightly greater than the background, which causes that image patch with defects are very similar to homogeneous image patch without defects and causes that the recognition of both classes can be very difficult.

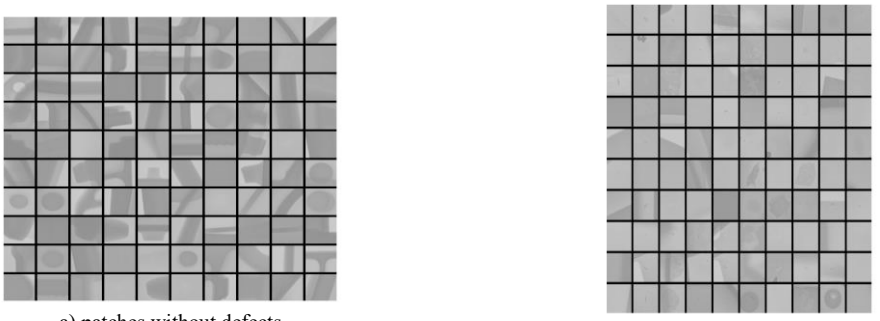
This paper attempts to pursue an approach with better performance on detection of casting defects in X-ray images from our InteCAST dataset. We collected 227 full size (3072 × 2400) X-ray images of supporting plates in aeromotors from a Chinese aerospace enterprise. 500 image patches(250 × 250) with defects and 500 image patches(250 × 250) without defects were cropped from all full size images. We compare and evaluate several methods, most of which are based on different feature engineering methods and machine learning models, including local binary patterns-SVM, Gabor-SGD, histogram of oriented gradient-random forest and combination among them. We believe that this paper possesses high reference value for similar task on computer aided detection.

The rest of the paper is organized as follows. First, the related works of this paper are explained in Section 2, and the assessment methods employed is described in Section 3. The obtained results are presented and discussed in Section 4. Finally, some conclusions and future work are reported in Section 5.



a) #A1033 b) #A1034

Fig. 1. Two examples of defects in real X-ray images of strut in aeromotor from InteCast dataset



a) patches without defects b) patches with defects

Fig. 2. Examples of patches containing defects (a) and no-defects (b) from our database

## 2. Related Works

Lots of approaches for computer aided detection of defects in casting X-ray images have been reported in the literatures during the past several years [4]. Background reconstruction applying a dictionary is a common method, in which an estimated and reconstructed background image without defects is subtracted from the original image leaving a new image with defects and noise [5-6]. Background reconstruction approaches were also applied to the task of welding defect detection successfully [7-9]. However, Background reconstruction methods tend to be very sensitive to the position of the defects and the noise in images [5]. There has been several released datasets for experimental purposes, such as the X-ray images dataset (GDXray) [10]. The performance on the GDXray welds series and castings series of several defect segmentation methods [11] are evaluated qualitatively. A comparative research [12] on various computer vision techniques for casting defects recognition has been reported, in which 32×32 pixels image patches were cropped from full size casting X-ray images in GDXray and were employed to assessment lots of different classifiers. The deep convolutional neural network has been reported that they possess the ability to learn features of an image and recognize the objects in an image [13, 24]. Several deep learning models were also evaluated [12], acquiring not more than 86% image patch classification accuracy, which demonstrates that scarcity of casting X-ray images with defects causes that deep learning approaches, which demand massive data to train an ideal model, perform not so well as traditional machine learning methods on casting defect detection task. The best solution, which also provides a reference for our task, is LBP feature based linear SVM classifier model.

### *2.1. Feature Engineering*

Feature engineering can be considered as a process of transforming raw images into features that possess the ability to better represent the underlying problem. Common features of images are mainly composed of local binary patterns (LBP), histograms of oriented gradients(HOG) and Gabor.

LBP (Local binary pattern), which was first proposed by T.Ojala for texture feature extraction, is used to describe local texture features of images and it has significant advantages such as rotation invariance and gray invariance [14]. A neighborhood centered on a pixel is the element to be calculated on. The pixels adjacent the centered pixel are marked as 0, if the intensity values of adjacent pixels is equal to or smaller than the intensity value of the center pixel , otherwise marked as 1. Binary digits are then applied to represent corresponding pixel in the LBP feature map. Several scales of neighbor domain and block sizes were tested in this paper and the best results on our task were obtained with 3 by 3 neighbor domain and 5 by 5 blocks.

HOG (histograms of oriented gradients) is a feature that is able to quickly extract the local gradient characteristics of an object. This feature engineering approach divides an image patch into blocks firstly, and then divides each block into cells, and then counts the gradient direction histogram of each cell. All gradient direction histograms of all cells were combined as the feature of an image [15]. Several orientations, size of per cell and sizes of per block were tested in this paper and the best results on our task were obtained with 9 orientations, 20 by 20 pixels per cell and 2 by 2 cells per block.

Gabor [16] is a feature that can be used to describe image texture information and is sensitive to image edge, being able to provide good direction and scale selection characteristics. Gabor feature is insensitive to light changes, being capable of providing good adaptability to light changes. The frequency and direction of Gabor filter are similar to human visual system. Its essence is 2D Gaussian-shaped bandpass filters, which are with dyadic treatment of the radial spatial frequency range and multiple orientations and especially suitable for texture representation and discrimination. Several scales and orientations were tested in this paper and the best results for our task were obtained with 4 orientations and 3 scales.

### *2.2. Machine Learning*

Traditional method requires manual programming of rules. However, machine learning is that a computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience [17].

The classifier models, which were all implemented with Python, tested in this paper include SGD-Classifier(sgd_clf)[18],SVM-Classifier(svc_clf,svcr_clf)[19],NaïveBayes-Classifier (GNB_clf,MNB_clf)[20], Random Forest–Classifier(rfc_clf) [22], AdaBoost–Classifier(adb_clf) [21], Gradient Boosting–Classifier(grb_clf)[23]. For the sake of clearer representation in the curve figures, all models were also marked as capital letter, as listed in Table 1.

Table 1. Classifiers and their corresponding code names.

| Model | sgd_clf | svc_clf | rfc_clf | GNB_clf | MNB_clf | svcr_clf | adb_clf | grb_clf |
|---|---|---|---|---|---|---|---|---|
| Code names | A | B | C | D | E | F | G | H |

The hyperparameters of each model are set as follows. Linear SVC(svc_clf), Gaussian Naïve Bayes Classifier (GNB_clf) and AdaBoost Classifier(adb_clf) were applied default hyperparameters . The random state of SGDClassifier is 42. The number of estimators , max depth, and random state of RandomForestClassifier are respectively 66 , 100 and 1234. Alpha of Multinomial Naïve Bayes Classifier is 0.01. The kernel of SVM Classifier is set as rbf, and SVM Classifier is set to output probability. The number of estimators of GradientBoostingClassifier is 200.

## 3. Assessment methods

As described in Section 1, the size of learning set, $\left\{X_{Learn}^{i}, y_{Learn}^{i}\right\}_{i=1}^{N_{Learn}}$ , is 800( $N_{Learn}=800$ , 400 samples from each class),and the testing set, $\left\{X_{Test}^{i}, y_{Test}^{i}\right\}_{i=1}^{N_{Test}}$ ,contains 200 samples( $N_{Test}=200$ , 100 samples from each class). X is defined as the set of all image patches, and y is defined as the set of the labels of corresponding elements in X. The learning set and the testing set are disjoint. Then we have the following equation:

$$X_{Learn} \cap X_{Test} = \varnothing \tag{1}$$

We define:

$$F_{Learn}^{i} = F\left(X_{Learn}^{i}\right) \tag{2}$$

$$F_{Test}^{i} = F\left(X_{Test}^{i}\right) \tag{3}$$

Where $F\left(x\right)$ is a processing operation representing feature engineering. After processing all image patches in X we get $\left\{F_{Learn}^{i}, y_{Learn}^{i}\right\}_{i=1}^{N_{Learn}}$ and $\left\{F_{Test}^{i}, y_{Test}^{i}\right\}_{i=1}^{N_{Test}}$ . $\left\{F_{Learn}^{i}, y_{Learn}^{i}\right\}_{i=1}^{N_{Learn}}$ is fed into a machine learning model, $H\left(x\right)$ , to train the model. After training,

$$H(F_{Learn}^{i}) \approx y_{Learn}^{i} \tag{4}$$

Testing set was applied on the trained model $H(x)$ to test whether the images can be classified correctly, which means that whether $H(F_{Test}^{i}) \approx y_{Test}^{i}$, where $F_{Test}^{i} = F\left(X_{Test}^{i}\right)$, is tested.

The confusion matrix, as shown in Fig.3, is adopted for quantitative assessment on the performance of a classifier.

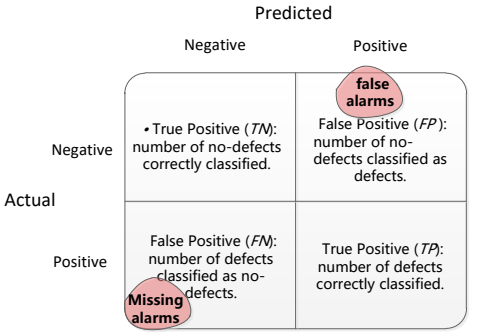



Fig. 3. An illustrated confusion matrix

The confusion matrix offers a lot of information, from which Precision (Pr), Recall (Re) and Accuracy (η) can be acquired.

$$Pr = \frac{TP}{TP + FP} \tag{5}$$

$$Re = \frac{TP}{TP + FN} \tag{6}$$

$$\eta = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}$$

$$F_1 = \frac{2}{\frac{1}{Pr} + \frac{1}{Re}} \tag{8}$$

Precision (Pr), Recall (Re), F1-score and Accuracy (η) of an ideal model all equal 1.0, which means that all defects are detected with no false alarms. In this paper, Precision versus Recall curve, the receiver operating characteristic (ROC) curve, precision, recall and accuracy of each model was reported.

## 4. Results and Discussion

Three feature engineering algorithms combined with eight kinds of machine learning models were tested applying the assessment method in Section 3 with 800 image patches as learning set and 200 image patches as testing set. All results of this paper have been shown in Table 2 and Fig. 4.

The higher Precision (Pr), Recall (Re), F1-score and Accuracy (η) of an approach is, the better performance of an approach is. The top three approaches are highlighted in red, green and blue, in Table 2. As for accuracy acquired by the all methods, the best three approaches, #24,16,19, 23, can be distinguished in Table 2 and Fig.6. #24,16,19,23, most of which are LBP based approaches, have a better accuracy with a range of 86% to 89%. #2,8,7,5,4,14,1,6,most of which are Gabor based methods, possess a low accuracy with a value less than 71%. From Fig.4, AUC(area under curve, e.g. accuracy) can be visualized through the curve of each method. Fig.5 (a-b) show that rfc_clf, adb_clf and grb_clf , all of which are ensemble learning model, are better models for Gabor feature, although the accuracy obtained by #3,7,8 are not more than 75%. Fig.4 (c-d) show that grb_clf, rfc_clf and adb_clf

are better models for HOG feature, with AUC being respectively 86%,84% and 83%. Fig.4 (e-f) show that grb_clf, rfc_clf and adb_clf are also better models the best model for LBP feature, with AUC being respectively 89%, 88% and 86%. Those results coincide with the theory that the ensemble learning model has better classification performance than the single model on the same data set.

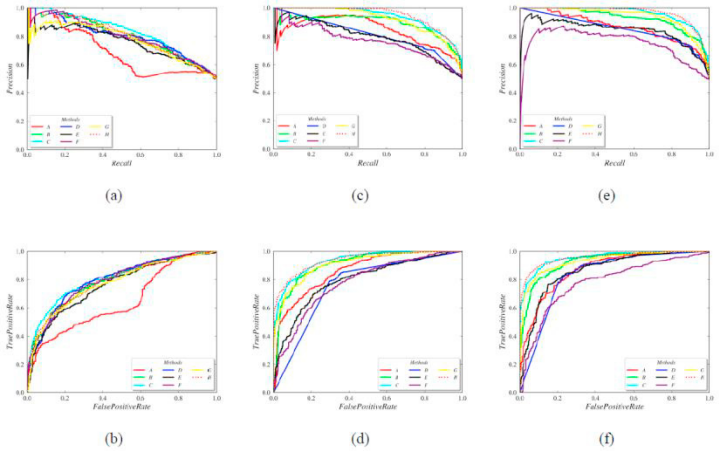



Fig. 4. Precision versus Recall curves and the receiver operating characteristic (ROC) curves of all methods. (a) and (b) are the curves of 8 Gabor feature based methods. (c) and (d) are the curves of 8 HOG feature based methods. (e) and (f) are the curves of 8 LBP feature based methods.

As for recall rates acquired by the all methods, different from accuracy and precision, the best three approaches are #6,9,17,18, the recall rates range of which is between 89% to 99% . From Fig.3 and equation (6), the higher a recall rate is, the less defects was classified as no-defects, which means that less defects will be missed. Thus, a higher recall rates is ideal in this paper, although more false alarms will be generated. Therefore, recall is the factor with high weight when picking the best approach for our task.

Table 2. Performances of all approaches in the task for computer aided defect detection on InteCast X-ray images dataset.

| # | **F** | **H** | TP | FP | FN | TN | Pr | Re | F1 | Accuracy | AUC |
|---|---|---|---|---|---|---|---|---|---|---|---|
| * | Ideal | Ideal | 400 | 0 | 0 | 400 | 1 | 1 | 1 | 1 | 1 |
| 1 | Gabor | sgd_clf | 327 | 73 | 230 | 170 | 0.69 | 0.42 | 0.52 | 0.62 | 0.62 |
| 2 | | svc_clf | 281 | 119 | 110 | 290 | 0.7 | 0.72 | 0.71 | 0.71 | 0.71 |
| 3 | | rfc_clf | 318 | 82 | 118 | 282 | 0.77 | 0.7 | 0.73 | 0.75 | 0.75 |
| 4 | | GNB_clf | 217 | 183 | 64 | 336 | 0.64 | 0.84 | 0.73 | 0.69 | 0.69 |
| 5 | | MNB_clf | 271 | 129 | 123 | 277 | 0.68 | 0.69 | 0.68 | 0.69 | 0.69 |
| 6 | | svcr_clf | 6 | 394 | 3 | 397 | 0.5 | ***0.99*** | 0.66 | 0.5 | 0.5 |
| 7 | | adb_clf | 282 | 118 | 118 | 282 | 0.7 | 0.7 | 0.7 | 0.71 | 0.71 |
| 8 | | grb_clf | 286 | 114 | 123 | 277 | 0.7 | 0.69 | 0.7 | 0.7 | 0.7 |
| 9 | HOG | sgd_clf | 233 | 167 | 39 | 361 | 0.68 | ***0.9*** | 0.77 | 0.74 | 0.74 |
| 10 | | svc_clf | 316 | 84 | 56 | 344 | 0.8 | 0.86 | 0.83 | 0.82 | 0.82 |
| 11 | | rfc_clf | 347 | 53 | 72 | 328 | 0.86 | 0.82 | 0.83 | 0.84 | 0.84 |
| 12 | | GNB_clf | 281 | 119 | 85 | 315 | 0.72 | 0.78 | 0.75 | 0.75 | 0.75 |
| 13 | | MNB_clf | 279 | 121 | 93 | 307 | 0.71 | 0.76 | 0.74 | 0.73 | 0.73 |
| 14 | | svcr_clf | 217 | 183 | 63 | 337 | 0.64 | 0.84 | 0.73 | 0.69 | 0.69 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 15 | | adb_clf | 334 | 66 | 69 | 331 | 0.83 | 0.82 | 0.83 | 0.83 | 0.83 |
| 16 | | grb_clf | 353 | 47 | 65 | 335 | ***0.87*** | 0.83 | 0.85 | ***0.86*** | ***0.86*** |
| 17 | | sgd_clf | 265 | 135 | 39 | 361 | 0.72 | ***0.9*** | 0.8 | 0.78 | 0.78 |
| 18 | | svc_clf | 310 | 90 | 42 | 358 | 0.79 | ***0.89*** | 0.84 | 0.83 | 0.83 |
| 19 | | rfc_clf | 364 | 36 | 60 | 340 | ***0.9*** | 0.85 | ***0.87*** | ***0.88*** | ***0.88*** |
| 20 | LBP | GNB_clf | 315 | 85 | 87 | 313 | 0.78 | 0.78 | 0.78 | 0.78 | 0.78 |
| 21 | | MNB_clf | 299 | 101 | 61 | 339 | 0.77 | 0.84 | 0.8 | 0.8 | 0.8 |
| 22 | | svcr_clf | 304 | 96 | 116 | 284 | 0.74 | 0.71 | 0.72 | 0.73 | 0.73 |
| 23 | | adb_clf | 346 | 54 | 56 | 344 | 0.86 | 0.86 | ***0.86*** | ***0.86*** | ***0.86*** |
| 24 | | grb_clf | 369 | 31 | 51 | 349 | ***0.91*** | 0.87 | ***0.89*** | ***0.89*** | ***0.89*** |

## 5. Conclusions

In this paper, a comparative assessment of 24 approaches, which can be potential to be applied in practice for computer aided X-ray image defect detection, was reported. Three feature engineering methods, Gabor, HOG and LBP, and corresponding eight machine learning models, SGD classifier, SVM classifier, Random Forest classifier, Naïve Bayes classifier, AdaBoost classifier, Gradient Boosting-classifier, were evaluated. Results in our experiments show that the best candidate for our task was acquired by LBP feature based Gradient Boosting-Classifier method, which obtains 89% of accuracy (with 91% precision and 87% recall) and is able to detect defects easily overlooked by naked eye, although it is tending to neglect defects that can be easily detected manually.

Although the LBP feature based Gradient Boosting-Classifier method in this paper is for future research, we plan to collect more X-ray images of strut in aeromotors to make a larger dataset, and to invest more time to develop an improved LBP feature engineering method and to tune hyperparameters of Gradient Boosting-Classifier. Then, we will run the improved approach on full size X-ray images applying sliding windows or region proposal strategy. If we are able to get more images, we will also try to train an end-to-end deep learning model [24].

## Acknowledgements

This work is financially supported by the Graduates' Innovation Fund, Huazhong University of Science and Technology (Grant No. 2019YGSCXCY015).

## References

[1] CIRP, 50 (2016) 512-517. Rajesh R, Khan J G. Defects, Causes and Their Remedies in Casting Process[J]. A Review, International Journal of Research in Advent Technology, 2014, 2(3): 817-827.
[2] Li X, Tso S K, Guan X P, et al. Improving automatic detection of defects in castings by applying wavelet technique[J]. IEEE Transactions on Industrial Electronics, 2006, 53(6): 1927-1934.
[3] Ghorai S, Mukherjee A, Gangadaran M, et al. Automatic defect detection on hot-rolled flat steel products[J]. IEEE Transactions on Instrumentation and Measurement, 2012, 62(3): 612-621.
[4] Mery D. Computer Vision for X-Ray Testing[J]. Switzerland: Springer International Publishing, 2015.
[5] Piccardi M. Background subtraction techniques: a review[C] 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583). IEEE, 2004, 4: 3099-3104.
[6] Rebuffel V, Sood S, Blakeley B. Defect detection method in digital radiography for porosity in magnesium castings[J]. Materials Evaluation, ECNDT, 2006.
[7] Nacereddine N, Zelmat M, Belaifa S S, et al. Weld defect detection in industrial radiography based digital image processing[J]. Transactions on Engineering Computing and Technology, 2005, 2: 145-148.
[8] Wang G, Liao T W. Automatic identification of different types of welding defects in radiographic images[J]. Ndt & E International, 2002, 35(8): 519-528.
[9] Kaftandjian V, Joly A, Odievre T. Automatic detection and characterization of aluminium weld defects: comparison between radiography, radioscopy and human interpretation[M]. Society of Manufacturing Engineers, 2000.

[10] Mery D, Riffo V, Zscherpel U, et al. GDXray: The database of X-ray images for nondestructive testing[J]. Journal of Nondestructive Evaluation, 2015, 34(4): 42.
[11] Mirzaei F, Faridafshin M, Movafeghi A, et al. Automated Defect Detection of Weldments and Castings using Canny, Sobel and Gaussian filter Edge Detectors: A Comparison Study[J]. Tehran, Iran, 2017.
[12] Mery D, Arteta C. Automatic defect recognition in x-ray testing using computer vision[C] 2017 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2017: 1026-1035.
[13] Makantasis K, Protopapadakis E, Doulamis A, et al. Deep convolutional neural networks for efficient vision based tunnel inspection[C] 2015 IEEE International Conference on Intelligent Computer Communication and Processing (ICCP). IEEE, 2015: 335-342.
[14] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, 2002.
[15] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 1, pages 886–893. IEEE, 2005.
[16] A.Kumar and G. Pang. Defect detection in textured materials using gabor filters. IEEE Trans. on Industry Applications, 38(2):425–440, 2002.
[17] Mitchell, Tom M. "Machine learning and data mining." Communications of the ACM 42.11 (1999).
[18] Vapnik V. Statistical Learning Theory. John Wiley&Sons[J]. Inc., New York, 1998.
[19] Cortes C, Vapnik V. Support vector machine[J]. Machine learning, 1995, 20(3): 273-297.
[20] McCallum A, Nigam K. A comparison of event models for naive bayes text classification[C] AAAI-98 workshop on learning for text categorization. 1998, 752(1): 41-48.
[21] Drucker H, Cortes C, Jackel L D, et al. Boosting and other ensemble methods[J]. Neural Computation, 1994, 6(6): 1289-1301.
[22] Breiman L. Random forests[J]. Machine learning, 2001, 45(1): 5-32.
[23] Friedman J H. Greedy function approximation: a gradient boosting machine[J]. Annals of statistics, 2001: 1189-1232.
[24] LeCun Y, Bengio Y, Hinton G. Deep learning[J]. nature, 2015, 521(7553): 436.