Deformable Feature Pyramid Network for Aluminum Profile Surface Defect Detection

Juan Wang*, ZhaoHui Meng

Computer and Information Department, HoHai University, NanJing, JiangSu, 211100, China

*Corresponding author's e-mail: ms_wangjuan@qq.com

Abstract. Surface defect detection is an important part in the process of aluminum production. The goal of a complete defect detection task is to realize the specific category and precise location of each defect in the image. Due to the similarity between surface defects and background, as well as the large diversity and difference in the appearance of the defects, this task is still challenging for applying this task in practice. In order to solve these problems, a deep learning-based aluminum profile defect detection network is proposed. In order to realize strong classification ability, the deep convolutional neural network is used to generate feature graphs at each stage. On this basis, the variable feature pyramid module (DFP) is proposed, and deformable convolution is added to the feature layer to make the feature layer have the ability to adapt to defect deformation. Based on these multi-layer features, the size of anchor box was customized by Kmeans clustering algorithm, and then the region extraction network (RPN) was used to generate the region of interest (ROIs). For each ROI, a detector, consisting of a classifier and a bounding box regressor, produces the final detection results. Finally, we used a defect detection dataset of aluminum profiles to train and evaluate our approach. Through two sets of ablation experiments, the effectiveness of the introduced module is proved. Finally, we reached 76.92 mAP in aluminum profile data set.

1. Introduction
Defect inspection is a crucial step to guarantee the quality of industrial production. Manual quality inspection has so many disadvantages. The automatic detection method based on machine vision has some improvements and has been applied in many fields. Zhang et al. [1] used the curvature filter and gaussian mixture model to detect the surface defects of the rail. Wang et al. [2] applied the template-based method to the surface defect detection of strip steel. Other methods based on manual characteristics have been used for defect detection in industrial applications (such as solar modules [3], metals [4], steel [5]), and have achieved good results in recent years. However, these traditional image processing techniques often require researchers to have professional knowledge, so it is affected by manual intervention and lacking of robust. In recent years, many methods based on deep learning [6] have been proved effective in visual tasks. Deep learning control the feature extraction by changing the CNN's width and depth. Unlike manual extraction characteristics, it can only get some single level features without hierarchical structure, which without enough generalization ability. Defect detection tasks usually take several forms: segmentation, classification, and detection. Due to the lack of specific data sets, many methods can only classify defects [7][8][9]. Although this method is simple, but too simplified, which unable to provide location information, and cannot provide more help for industrial quality inspection. Mask r-cnn [10], a segmenter based on deep learning had a more specific
description for the shape of the defect. However, segmentation tasks consumes a large amount of computing resources and cannot meet the real-time check requirements of industrial detection. In addition, it is costly to set up a large defect - partitioned data sets. Therefore, we tried to build an end-to-end automated defect detection system (ADD). ADD can provide a bounding box with class scores to accurately classify and locate defects. In this task, we encountered the following challenges:

![Figure 1. Examples of flaw patterns. (a) spray, which is very similar to the background. (b) paint bubble, narrow and long. (c) paint bubble, narrow and long. (d) orange peel, which including the entire aluminum profile surface. (e) scratch, irregular defect form.](image)

First, the detection system needs a strong classification ability to distinguish the defects similar to the background (Fig1 (a)). Therefore, we need a deep convolutional neural network as the backbone network. While it is easy to overfit driving a large network on a small data set. However, industrial data sets costs high, it is expensive to obtaining large amounts of data. Second, aluminum profile defects are complex and changeable, lacking fixed shape features. For example, there are narrow and long defects in Fig1 (b), and the square defects in Fig1 (c). Besides, defects scale vary largely. Some defects account for the entire surface of the aluminum profile, which is shown in Fig 1(d). While the defects in Fig1 (c) are less than one percent of the aluminum profile surface. Sometimes, the shape of the defects is irregular, just like Fig (e). Finally, the deep learning framework strengthen the network classification ability, which also raises some problems. For example, many object detection systems [11] are classified and located based on the last feature map. We know that during the convolution process of a deep convolutional neural network, the feature map is scaled. If the feature is small before scaling, then it is not discriminating after scaling. In addition, in some common two-stage object detection frameworks (such as Fast-RCNN, Faster-RCNN, RFCN [11] [12] [13]), the anchor box ratio are defined beforehand. In general, a set of anchor boxes with a ratio of {1: 1, 1: 2, 2: 1} can cover most object sizes in natural scenes. Many defect forms are extreme for defects in industrial aluminum profiles, such as the Fig1 (b), so the predefined anchor size cannot accurately fit such defects, the localization of the defects will have bias, which affecting the effect of defect detection eventually.

Based on the above problems, we proposed a deformable feature pyramid network (DFPN) to try to improve the above problems. First of all, in the current study of transfer learning [14], we can perform preprocessing on ImageNet. Be inspired by [15], using the pre-training weights that are pre-trained on the COCO object detection dataset will have better performance than the classification-based dataset when dealing with object detection problems. Besides, deformable convolution is introduced, which can automatically calculate the offset of each point in the process of convolution calculation, so as to extract features from the most suitable place for convolution and avoid learning unnecessary background features. In addition, the deformable convolution is fused with FPN [16], and
the multi-layer feature map obtained by the convolution network is extracted for feature fusion. This has two advantages. First, the low-level features are fused with high-level information after convolution and upsampling operations. In the convolutional neural network, high-level features have strong semantic information, and low-level features have structural information. Therefore, combining the high-level and low-level information can enhance the expression ability of features. Second, we scattered the location information of the features information generate and extract by proposal bounding boxes to each layer of the feature pyramid, which can increase the feature mapping resolution of small targets and is also good for the final prediction.

In summary, the main contributions of this article are:

- Introduced a deep learning-based surface defect detection method that integrates ResNet and RPN for accurate defect classification and location.
- A deformable feature pyramid network (DFPN) is proposed. Compared with other fusion methods, DFP can better adapt to feature shapes.
- The method uses Kmeans clustering algorithm to customize the anchor box, which can better fit aluminium’s defect bounding boxes.

2. Related work

In recent years, surface defect detection methods can be divided into traditional machine vision-based methods and deep learning-based computer vision methods.

2.1. Traditional machine vision-based methods

Traditional machine vision inspection methods mainly rely on cameras, external light sources, lenses and other equipment to collect defect images of target products, and then using machine vision algorithm libraries based on the principles of digital image processing and analysis such as OpenCV, Halcon, VisionpPro to detect defects through a series of operations such as segmentation, morphological processing, edge detection, feature extraction and so on.

2.2. Deep learning-based computer vision methods

The Surface defect detection methods based on deep learning can be divided into image-level defect classification, region-level defect detection and pixel-level defect segmentation according to the detection task. Segmentation refers to the classification of each pixel in an image. Wang et al. [17] combined wavelet edge detection and multi-scale structured forest to improve the accuracy of surface crack segmentation. Zhang et al. [18] achieve automatic segmentation of nanoparticles based on a U-Net convolutional neural network. Classification is a common defect detection task. Zhang et al. [19] extracted the feature vectors of steel strip defects based on statistical methods and spectral measurements, and built an online classification system based on support vector machines. Detection including two tasks, which is the localization and classification of defects in the image, which is usually more difficult than classification. Chang et al. [20] implemented defect detection of industrial CT images based on Faster RCNN, Liong et al. [21] built automatic defect detection and segmentation system for leather based on Mask RCNN. However, none of these methods analyzes aluminum profile defects or do the algorithmic research and discussion of its variable shapes, large scale variance and small flaws. While these are their common feature in the process of industrial aluminum profile flaw detection.

3. Defect detection network

This section describes the automatic defect detection network (ADD) in detail, as shown in Fig2. It is mainly divided into 3 parts: feature extraction, feature fusion, and proposal region generation RPN. Input a picture of any size, which through some operations performed by CNN such as convolution pooling to generate feature maps at each stage. In the last layer of the convolutional network, we add an offset in convolution of the last block of backbone so that the corresponding feature map has the ability to adapt to the deformation object. Then we use FPN to fuse the multi-layer features, which is
The region proposal network (RPN) [10] is used to generate region of interests (ROIs) on DFP’s features. Finally, the DFP features corresponding to each ROI are converted into fixed-length features and sent to two fully connected layers through RoI Pooling. One of the fully connected layer is a (C + 1) defect classification layer (cls), and the other is a bounding box regression layer (loc).

3.1. Baseline ConvNet Architecture

CNNs are widely used to extract features according to the characteristics of objects, which can be learned by superimposing multiple convolutional layers and pooling layers. In this paper, we choose resnet101 as the backbone network. ResNet has several attractive advantages. On the one hand, ResNet can obtain the highest accuracy with very few parameters compared with CNN of the same order of sequential pipeline structure. On the other hand, ResNet is modularized, which is helpful to subsequent feature fusion. ResNet introduces a residual network structure. In the residual network, the network is not directly fitted to the original mapping, but the residual mapping. Suppose the original mapping is $H(x)$, and the mapping of the residual network fitting is: $F(x) = H(x)$. It has two layers. As follows. $F = W_2 \sigma(W_1x)$. Where $\sigma$ represents the non-linear function ReLU, $x$ is the input. Then through a shortcut structure and the second ReLU to get the output $y$. $y = F(x, [W_i]) + x$.

3.2. Deformable feature pyramid network

In the object detection task, many frameworks [11] [12] only use high-level features to extract proposal regions, so the semantic information of the underlying features cannot be used. The feature pyramid module [16], which add a top-down semantic information transmission path low-level feature map also obtain some deep semantic information. However, this methods ignores that how to do deal with irregular defects.
Therefore, we introduce a deformable convolution, which adds an offset to each point on the receptive field. The size of the offset is learned. The receptive field after offset is no longer a square, but matches the actual shape of the object. The intuitive result is shown in Fig3. It can be seen that the receptive field of the traditional convolution is always a fixed square in Fig3 (a), while the deform convolution in Fig3 (b) is basically concentrated on the location of defects after two convolutions because each position of the receptive field has an offset. The traditional convolution structure can be defined as Equations (1), where $p_n$ is the offset of each point in the convolution output corresponding to each position in the perceptive field, which is an integer. After using the new deformable convolution, an offset $\Delta p_n$ is added to Equation (1). This new offset is derived from another convolution, it is generally a decimal, which is shown in Equation (2).

$$y(p_0) = \sum_{p_n \in \mathbb{R}} w(p_0) \cdot x(p_0 + p_n)$$

(1)

$$y(p_0) = \sum_{p_n \in \mathbb{R}} w(p_0) \cdot x(p_0 + p_n + \Delta p_n)$$

(2)

The value position of $x(p_0 + p_n + \Delta p_n)$ in Equations (2) is not an integer and does not correspond to the actual points on the feature map, so it must be obtained by interpolation. Here we use the bilinear interpolation method. So how do we fuse the deformable convolution and our feature extraction network? As shown in Fig3, we changed the last 3 convolution layers of the last blocks of resnet101 into deformable convolution, and was fused the resulting feature map with FPN. Why we do not use earliest layers? Because the offset learning requires certain semantic features and the experiments shows that it has better results by using the last 3 layers. Besides, it cost less. Now, our feature extraction network DFP not only has the capability of multi-scale defect feature extraction, but also can adaptively learn the feature extraction of irregular defects.

3.3. Extract region proposal
In this paper, RPN slides on the four-layer feature maps to extract region proposals. In the original RPN design, there were three types of anchor boxes with an aspect ratio of \{1: 1, 1: 2, 2: 1\}. By sampling at different sizes, the size of most objects can be basically fitted. However, in the detection of aluminum profile defects, there are many defects of extreme size, as shown in Fig2. In this way, the original ratio of anchor box cannot be adapted to the aluminum profile defect detection in the practical application. Therefore, based on the inspiration of [22], this paper introduces the Kmeans clustering algorithm to analyze the real distribution of anchor boxes in the training set, so as to obtain the custom anchor size. With this way of customizing the anchor box, we can more accurately fit real samples, accelerate model convergence, and improve model detection accuracy.

4. Experiment
This article uses pytorch-1.1.0 as the experimental framework. The purpose of the experiment is mainly to verify the validity of the added module. Therefore, this article mainly performed two ablation experiments on the data set, which is an aluminum profile defect data set provided by the Alibaba platform. It including 3005 pictures, which are divided into ten categories, including rubbing, mottled, bottom leakage, non-conductive, orange peel, spray, paint bubbles, Pitting, dirty spots, corner leaking. In this paper, 2404 samples are selected as the training set and 601 samples are used as the test set. Due to the limited data set, this article uses a coco pre-trained model and fine-tunes this model using the aluminum profile defect data set. This article uses the average accuracy (mAP) to evaluate the quality of the test results. $\text{Precision} = \frac{TP}{TP + FP}$, $\text{Recall} = \frac{TP}{TP + FN}$, $AP = \frac{\text{Precision} + \text{Recall}}{2}$.

In the above Equations, TP, FP, and FN respectively present the number of correctly classified positive examples, that is, the number of instances (samples) that are actually positive and classified
by the classifier as positive examples, the number of incorrectly classified as positive examples, that is, the number of instances that are actually negative but classified as positive by the classifier, the number of instances that were incorrectly classified as negative, that is, the number of instances that are actually positive but classified by the classifier as negative. Calculate the average accuracy (mAP) to evaluate the overall performance, which is the average of all classes of APs. Considering the calculation cost, all pictures are uniformly resized to 1280 × 960.

4.1. Deformable Feature Pyramid Module
Table 1 shows the results of detecting the aluminum profile defect detection data set when using resnet101 as the backbone network and faster-rcnn as the object detection framework, it is the baseline, and the ablation experiment after adding FPN and DFPN. It can be seen that when using FPN (introducing multi-level features for object detection) and using our module DFP (introducing deformable convolution to feature maps for detection), mAP have been improved in the aluminum profile defect dataset to some extent.

Table 1. The results of DFP module on aluminum profile data set.

|          | Baseline | FPN     | DFP    |
|----------|----------|---------|--------|
| mAP      | 72.5     | 75.09   | 76.32  |

4.2. K-means clustering algorithm custom anchor box
Table 2 shows the improvement results of aluminum profile defects when adding the k-means clustering algorithm to customize the anchor boxes based on Table 1.

Table 2. The results of DFP+k-means models on aluminum profile data set.

|              | DFP       | DFP+k-means |
|--------------|-----------|-------------|
| mAP          | 76.32     | 76.92       |

From the above experiments, it can be seen that our network improved 1.83% over FPN on aluminum profiles data set of without extra training data, compute, but only by improving the algorithm, indicating the effectiveness of the introduction module.

5. Conclusion
In this paper, by introducing deformations feature pyramid module (DFP) and k-means clustering algorithm to customize anchor box, which reached 76.92% mAP finally. There are still many improvements can be made in this paper. For example, in the aluminum profile defect data set, the number of defect object is very small, so we can do some screening for the candidate box generated by RPN, which can greatly improve the detection speed.

References
[1] Zhang, H., Jin, X.T, Wu, Q. M. J., Wang, Y., Yang, Y. (2018). Automatic visual detection system of railway surface defects with curvature filter and improved gaussian mixture model. IEEE Transactions on Instrumentation and Measurement, 67(7), 1-16.
[2] Wang, H.Y., Zhang, J., Tian, Y., Chen, H. Y., Liu, K. (2018). A simple guidance template-based defect detection method for strip steel surfaces. IEEE Transactions on Industrial Informatics, PP(99), 1-1.
[3] Tsai, D. M., Wu, S. C., Chiu, W. Y. (2013) Defect detection in solar modules using ica basis images. IEEE Transactions on Industrial Informatics, 9(1), 122-131.
[4] Wang, J. Li, Q. Gan, J. Yu, H. Yang, X. (2019) Surface defect detection via entity sparsity pursuit with intrinsic priors. IEEE Transactions on Industrial Informatics, PP(99), 1-1.
[5] Qiwu, L., Yichuang, S., Pengcheng, L., Oluymi, S., Lu, T., Yigang, H. (2018). Generalized completed local binary patterns for time-efficient steel surface defect classification. IEEE Transactions on Instrumentation and Measurement, 1-13.

[6] Lecun, Y., Bengio, Y., Hinton, G. (2015). Deep learning. . 521(7553), 436.

[7] Li, Y., Yi, G. Jiang, M. (2017) An end-to-end steel strip surface defects recognition system based on convolutional neural networks. Steel Research Int. 88(2), 176-187.

[8] Zhou, S., Chen, Y. Zhang, D. (2017) Classification of surface defects on steel sheet using convolutional neural networks. Material in Tehnologije, 51(1), 123-131.

[9] Natarajan, V., Hung, T. Y., Vaikundam, S. Chia, L. T. (2017) Convolutional networks for voting-based anomaly classification in metal surface inspection. In: IEEE International Conference on Industrial Technology. Toronto. pp. 986-991.

[10] Kaiming He, Georgia Gkioxari, Piotr Dollar, & Ross Girshick. (2017). Mask R-CNN. In: IEEE International Conference on Computer Vision. Venice. pp. 2980–2988.

[11] Girshick R. (2015) Fast R-CNN. In: IEEE International Conference on Computer Vision. Santiago .pp. 1440–1448.

[12] Ren S., He K., Girshick R.,Sun J. (2017) Faster RCNN: Towards real time object detection with region proposal networks. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(6):1137–1149.

[13] Dai J., Li Y., He K., Sun J. (2016) RFCN: object detection via region based fully convolutional networks. In: Neural Information Processing Systems Conference. Barcelona. pp. 379–387.

[14] J. Deng, W. Dong, and R. Socher , R., Li, L. J. , & Li, F. F . (2009). ImageNet: a Large-Scale Hierarchical Image Database. In: Computer Society Conference on Computer Vision and Pattern Recognition .Miami, Florida. pp. 248-255.

[15] Yosinski, J., Clune, J., Bengio, Y., Lipson, H. .(2014). How transferable are features in deep neural networks?. In: Proc. Neural Information Processing Systems Conference. Montréal. pp. 3320-3328

[16] T.-Y. Lin, P. Dollar, R. B. Girshick, K. He, B. Hariharan , S. J. Belongie. (2017). Feature pyramid networks for object detection. In: The IEEE Conference on Computer Vision and Pattern Recognition. Hawaii .pp.4.

[17] Wang, S., Wu X, Zhang Y.H., Chen, Q. (2018).Surface Crack Segmentation Based on Multi-Scale Wavelet Transform and Structured Forest. Acta Optica Sinica, 38(8): 0815024-1.

[18] Zhang, F., Wu, Y., Xiao, Z.T. Geng, L. Wu, J. Liu, Y.b. Wang, W. (2019) Nanoparticle Segmentation Based on U-Net Convolutional Neural Network. Laser & Optoelectronics Progress, 056(006): 129-135.

[19] Zhang, X.W., Ding, Y.Q., Lv, Y.Y. Shi, A.Y. Liang, R.Y. (2011) A vision inspection system for the surface defects of strongly reflected metal based on multi-class SVM. Expert Systems with Applications, 38(5): 5930–5939.

[20] Chang, H.T., Gou, J.N., Li, X.M. (2018) Application of Faster R-CNN in image defect detection of industrial CT. Journal of Image and Graphics, 23(7): 1061-1071.

[21] Liong, S. T ., Gan, Y. S ., Huang, Y. C ., Yuan, C. A ., Chang, H. C . (2019). Automatic defect segmentation on leather with deep learning. https://arxiv.org/abs/1903.12139?context=cs.

[22] Redmon, J., Farhadi, A.. (2018). Yolov3: an incremental improvement. http://xueshu.baidu.com/usercenter/paper/show?paperid=e02671f7b0527c6ecee43ce8bd7918b6&site=xueshu_se.