A GENERATIVE ADVERSARIAL NETWORK BASED FRAMEWORK FOR UNSUPERVISED VISUAL SURFACE INSPECTION

Wei Zhai *Jiang Zhu* *Yang Cao* *Zengfu Wang*

Department of Automation, University of Science and Technology of China
{wzhai056, zj130129}@mail.ustc.edu.cn, {forrest, zfwang}@ustc.edu.cn

ABSTRACT

Visual surface inspection is a challenging task due to the highly inconsistent appearance of the target surfaces and the abnormal regions. Most of the state-of-the-art methods are highly dependent on the labelled training samples, which are difficult to collect in practical industrial applications. To address this problem, we propose a generative adversarial network based framework for unsupervised surface inspection. The generative adversarial network is trained to generate the fake images analogous to the normal surface images. It implies that a well-trained GAN indeed learns a good representation of the normal surface images in a latent feature space. And consequently, the discriminator of GAN can serve as a naturally one-class classifier. We use the first three conventional layer of the discriminator as the feature extractor, whose response is sensitive to the abnormal regions. Particularly, a multi-scale fusion strategy is adopted to fuse the responses of the three convolution layers and thus improve the segmentation performance of abnormal detection. Various experimental results demonstrate the effectiveness of our proposed method.

Index Terms— visual surface inspection, unsupervised learning, generative adversarial networks, multi-scale fusion.

1. INTRODUCTION

Visual surface inspection, which aims to detect the abnormal regions in the surface of the workpiece by using computer vision techniques, draws a lot of attention due to its intense demands in industrial applications. Visual surface inspection is a challenging task due to severe image noises, large variation in the target surface, and strong diversity of abnormal regions.

Traditional visual surface inspection methods usually apply texture analysis techniques to detect the abnormal regions, such as structural-based method [1], statistical-based method [2] and filter-based method [3]. However, the performance of these kinds of methods drops significantly when they are applied to the different surfaces or abnormal regions.

To overcome this problem, learning based methods [4–6], e.g. Support vector machine (SVM), have been applied to learn a mapping between low-level features and abnormal regions. Recently, deep learning based methods have achieved great improvement on image related tasks such as object recognition [7–9]. Ren et al. [6] propose to use the weights of a pretrained deep CNN as feature extractor to detect the abnormal regions and achieve good performance on several industrial datasets. However, in practical industrial applications, it is difficult to collect a sufficient number of labelled training samples, especially for the samples with labelled abnormal regions.

In this paper, we propose an unsupervised learning based framework for visual surface inspection, in which no labelled abnormal samples are required. Our goal is to learn a mapping from the normal surface images to the latent feature space, and then to apply the mapping to the unseen surface images for abnormal inspection. To achieve this goal, we need to overcome the following challenges. (1). Develop a unified framework for various surfaces. Different workpieces have various surfaces. Even for a single surface, the appearances of different regions in the surface may be different. The proposed framework should be adapted to different surfaces inspection tasks. (2). How to extract effective features to describe normal surface texture. Previous works usually use hand-crafted features, which have low representation capability for surface inspection. New feature extractor specially designed for different surfaces is needed. (3). Existing surface inspection datasets are not sufficient to support and evaluate deep learning based research. The public dataset only contains several hundred of samples. It is easy to make deep neural networks over-fitting with such little samples.

Considering these challenges, we propose a generative adversarial network based framework for visual unsupervised surface inspection. In our proposed framework, the generative model (generator) learns to generate fake images analogous to the normal surface images, while the discriminative model (discriminator) learns to determine whether an image is from the sample distribution. It implies that a well-trained generative adversarial network indeed learns a representation of the normal samples distribution in a latent feature space. Therefore, we use the first three convolutional layers of the discriminator as a feature extractor, and use it to map the target surface images into the latent feature space. The da-

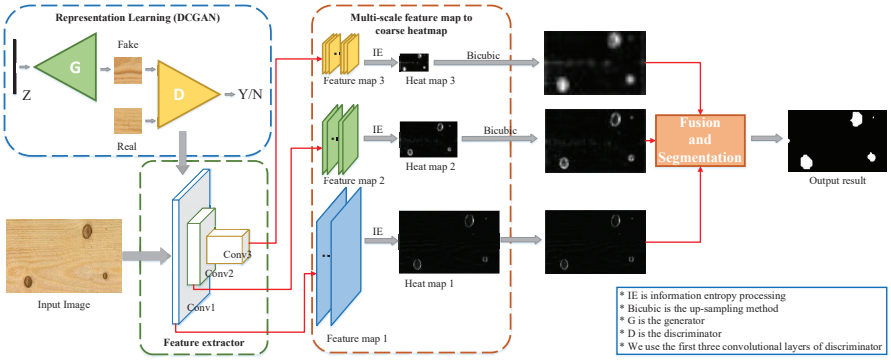


Fig. 1. The overall architecture of our proposed method for visual surface inspection.

ta points in the latent feature space, which are not approximate to the normal samples distribution, will be classified as abnormal. To further improve the performance of abnormal detection, a multi-scale fusion strategy is adopted to fuse the multi-scale feature maps into the finest scale. An adaptive threshold segmentation method is then applied to determine the final abnormal regions. Fig. 1 illustrates the details of our proposed method.

Our proposed framework has the following advantages:

1) A generative adversarial network based framework for visual surface inspection is presented. Since there is no need of labelled abnormal samples, it is much more suitable for practical application in industrial occasions.

2) Our GAN model learns the data distribution of the normal surfaces in the latent feature space. Therefore, the convolutional layers of the discriminator can be used as an extractor for surface specific features, which are more effective and robust than handcrafted feature extractor.

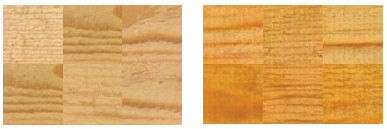
3) A multi-scale fusion strategy is adopted. It not only reduces the possibility of missing detection, but also improves the accuracy of the location of abnormal regions.

2. PROPOSED METHOD

2.1. DCGAN for feature representation learning

In recent years, generative adversarial networks have emerged as a powerful generative model and have been widely studied in previous works [10–13]. Surface inspection is substantially an one-class problem. A well-trained generative model indeed learns a representation of the data distribution in a latent feature space. The data points in the latent feature space, which are not approximate to the data distribution will be classified as fake.

DCGAN [12] is a strong candidate for unsupervised learning, thus we use it to learn a representation of the normal samples distribution. The generator G learns a distribution p_g over normal surface data x via a mapping $G(z)$ of samples z,



(a) (b)

Fig. 2. (a) Real images. (b) Fake images generated by GAN.

one-dimensional ($1D$) vectors of uniformly distributed input noise sampled from latent space Z, to two-dimensional ($2D$) images in the image space manifold χ which is populated by normal surface samples. Fig. 2 shows real and fake samples generated by DCGAN on Wood Defect Database (WOOD) [14]. G and D are simultaneously optimized through the following two-player minimax game with value function:

$$\min_G \max_D V(G,D) = \mathbb{E}_{\mathbf{x}\sim p_{data}(\mathbf{x})}[log(D(\mathbf{x}))] + \mathbb{E}_{\mathbf{z}\sim p_z(\mathbf{z})}[log(1\text{-}D(G(\mathbf{z})))]. \quad (1)$$

G is trained to generate fake images analogous to the normal surface samples and fool D. D is simultaneously trained to estimate the probability that a sample is from the training normal surface data rather than G.

2.2. Generation of inspection map with GAN's discriminator

We propose to use the weights of Neural Networks as a feature extractor to represent the texture structures of the normal surfaces. Intuitively, the discriminator of a well-trained GAN is a naturally one-class classifier. This leads to the idea of training the GAN with sufficient normal surfaces and using the weighs of discriminator as the feature extractor. A typical example is shown in Fig. 3, where the discriminator and generator are well-trained using normal wood surface images. Here we use information entropy [15] as a metric to represent

the output of the second convolutional layer in the discriminator. As can be seen, the response is insensitive to the normal regions but varies drastically across the abnormal regions. As a result, we use the first three convolutional layers of the discriminator as the desired feature extractor in this paper.

(a) (c) (b)

Fig. 3. (a) Input image. (b) Feature maps of $conv\_2$ in discriminator. (c) Inspection map calculated by information entropy.

Feature maps of the $k\text{-}th$ convolutional layer in discriminator is denoted as a matrix F. The feature vector in the position coordinate (x, y) can be represented as following:

$$f_{(x,y)} = [F(x, y, 1), F(x, y, 2)...F(x, y, m_k)]^T. \quad (2)$$

Here $f_{(x,y)}$ is the feature vector at (x, y), m_k is the number of channels in the $k\text{-}th$ convolutional layer.

Thus, the information entropy of feature vectors is expressed as following:

$$H_{(x,y)} = \sum_{m_k}^{i} f^i_{(x,y)} * log f^i_{(x,y)}. \quad (3)$$

Here, the $H_{(x,y)}$ is the calculated value of information entropy and we refer to H as a coarse inspection map, which reveals the abnormal regions of input image in Fig. $3(c)$.

2.3. Multi-scale fusion and abnormal segmentation

One of the challenges for surface detection based on texture analysis methods is that textures do not have scale invariance. Despite impressive surface detection results achieved by previous works [1,6,14], how to exploit a more effective strategy to address the aforementioned challenge? - largely remains to be studied. In this paper, we adopted a multi-scale heat map fusion strategy to deal with the challenge.

The first three convolutional layers of discriminator generate perceptually multi-level features and convolutional features in discriminator gradually change from low-level gradient feature to high-level structure feature with the increasing network depth. From such observations, we first resize the inspection maps produced by each convolutional layer to the same size, then we apply a weight average method for the fusion. The fusion process expresses as following:

$$R = \alpha H_1 + \beta H_2 + \gamma H_3. \quad (4)$$

Denote R as the fusion result and α, β, γ $(\alpha + \beta + \gamma = 1)$ are assigned weight values. The weight values are varied for different kinds of surfaces. H_k $(k = 1, 2, 3)$ is the heat map produced by the $k\text{-}th$ convolutional layer.

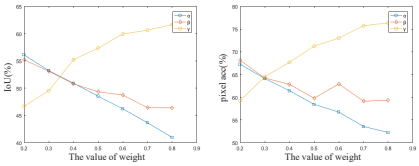


Fig. 4. The influence of changing weight values α, β and γ.

As for the weight setting, we preform the test, in which we change one of the weights while keep the other two weights fixed. As shown in Fig. 4, we observe that the weight values α and β are negatively related to segmentation results, but the weight values γ is positively related to segmentation results. Compared with $conv\_1$ and $conv\_2$, $conv\_3$ has larger receptive field and captures more contextual information. Therefore, we set a higher weight for $conv\_3$ layer.

After obtaining the final fusion heat map, we use Otsu's method [16] to binarize and acquire the segmentation results.

3. EXPERIMENTAL RESULTS

In this section, we compared our method with three baseline methods as the following: (1). Texture segmentation method proposed in [17] (ICPR 2010); (2). Object proposals method proposed in [18] (ICCV 2013); (3). Automated surface inspection method proposed in [6] (TCYB 2017).

Experimental data: Two datasets were used for training and testing respectively: (1). Wood Defect Database (WOOD) [14]; (2). Road Crack Database (CRACK) [19]. For the training of our GAN on each database, we extracted 50000 normal surface patches with size of 96×96 and used these samples as training data. To compare our method with TCYB 2017 [6], 50000 normal and 50000 abnormal surface patches from two datasets [14] were sampled to train their CNN using the strategy recommended by Ren et al. [6]. For the testing, we used 48 abnormal surface samples from WOOD and CRACK database, respectively. Both of the two datasets don't contain pixel-level segmentation labels, therefore we have made pixel-level segmentation labels of testing samples manually. $Groundtruth$ (GT) in Fig. 5 shows the results of our manual annotation.

Implementation details: For the training of our GAN, we set mini-batch size to 64 and the rest parameter setting was same as DCGAN [12]. The significance of our multi-scale fusion strategy can be seen in Fig. 6. To get the best fusion result, we set $\alpha = 0.2$, $\beta = 0.3$ and $\gamma = 0.5$, respectively. For

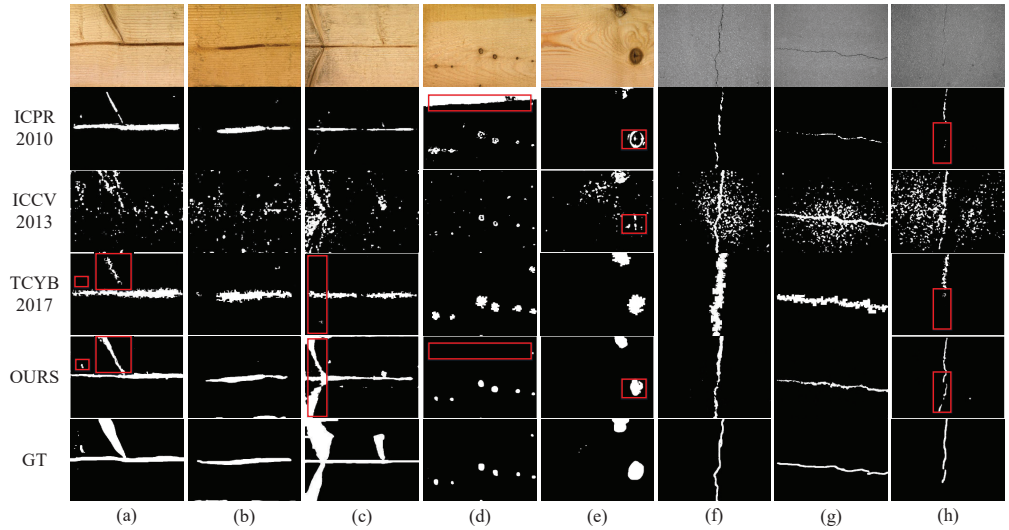


Fig. 5. The comparisons of visual surface inspection results with different methods.

better contrast, we also used Otsus method [16] to binarize the results of three baseline methods. Our method was implemented with tensorflow [20] and a TITAN Xp GPU was used for the training and testing.

Evaluation indexes: Two evaluation indexes in [21] are adopted to compare and analyze experimental results: (1). Intersection over union (IoU); (2). Pixel accuracy (pixel acc).

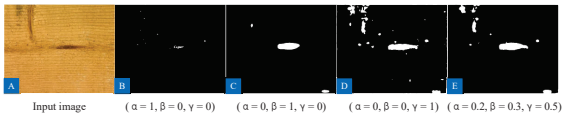


Fig. 6. The inspection maps generated with different weight values.

Table. 1 shows quantitative results of the evaluation indexes IoU and pixel acc of the four methods on the WOOD and CRACK dataset. As we can see in Table. 1, our method achieved higher IoU and pixel acc than the three baseline methods, e.g. ICPR 2010 [17], ICCV 2013 [18] and TCYB 2017 [6].

Table 1. Quantitative comparisons. (IoU(%) / pixel acc(%))

	ICPR 2010 [17]	ICCV 2013 [18]	TCYB 2017 [6]	Ours
WOOD	25.02 / 41.51	33.89 / 47.80	57.16 / 70.20	**63.90 / 79.85**
CRACK	29.97 / 42.33	18.48 / 31.10	29.36 / 43.05	**42.19 / 58.82**

Fig. 5 shows the visual comparisons of different methods for surface inspection. Although TCYB 2017 [6] located the abnormal surface regions more accurately than the other two baseline methods. Yet, our method achieve a better locating result. More importantly, our method only needs normal surface samples for training process. While the generic deep-learning-based approach TCYB 2017 [6] needs both normal and abnormal surface training samples. Lack of abnormal surface training samples will prevent it from training.

As shown in Fig. 5(a) (c) (h), our method detected almost all of the abnormal surface regions compared with three baseline methods. Three baseline methods appeared missed detection of some defect regions. And our method detected a challenging edge abnormal in Fig. 5(b). It demonstrated our method learned a more powerful surface texture feature extractor. It can be seen in Fig. 5(d), the abrupt transition of image color affected detection results of ICPR 2010 [17]. Yet, our method is more steady. In general, our method achieved better performance in terms of accuracy and robustness than the compared methods.

4. CONCLUSION

In this paper, we proposed a generative adversarial network based framework for unsupervised visual surface inspection. First of all, the normal surface samples were used to train D-CGAN for feature representation learning. Then, we used the first three convolutional layers of the discriminator as feature extractor and coarse inspection maps are obtained with information entropy processing. In order to get a better result, a multi-scale fusion strategy is adopted to fuse these coarse inspection maps. Finally, the abnormal regions segmentation results are produced with Otsu's method [16]. Comparing with three baseline methods, experimental results demonstrated the effectiveness and robustness of our method.

5. REFERENCES

[1] W. Wen and A. Xia, "Verifying edges for visual inspection purposes," *Pattern Recognition Letters*, vol. 20, no. 3, pp. 315–328, mar 1999.

[2] C.-W. Kim and A.J. Koivo, "Hierarchical classification of surface defects on dusty wood boards," in *[1990] Proceedings. 10th International Conference on Pattern Recognition*. 1994, IEEE Comput. Soc. Press.

[3] Jonathan G. Campbell, "Automatic visual inspection of woven textiles using a two-stage defect detector," *Optical Engineering*, vol. 37, no. 9, pp. 2536, sep 1998.

[4] Xue Wu Zhang, Fang Gong, and Li Zhong Xu, "Inspection of surface defects in copper strip using multivariate statistical approach and SVM," *International Journal of Computer Applications in Technology*, vol. 43, no. 1, pp. 44, 2012.

[5] Sarah M. Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning," *Pattern Recognition*, vol. 58, pp. 121–134, oct 2016.

[6] Ruoxu Ren, Terence Hung, and Kay Chen Tan, "A generic deep-learning-based approach for automated surface inspection," *IEEE Transactions on Cybernetics*, pp. 1–12, 2017.

[7] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton, "Imagenet classification with deep convolutional neural networks," in *International Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[8] Karen Simonyan and Andrew Zisserman, "Very deep convolutional networks for large-scale image recognition," *Computer Science*, 2014.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[10] Ian J. Goodfellow, Jean Pougetabadie, Mehdi Mirza, Bing Xu, David Wardefarley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, "Generative adversarial networks," *Advances in Neural Information Processing Systems*, vol. 3, pp. 2672–2680, 2014.

[11] Mehdi Mirza and Simon Osindero, "Conditional generative adversarial nets," *Computer Science*, pp. 2672–2680, 2014.

[12] Alec Radford, Luke Metz, and Soumith Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *Computer Science*, 2015.

[13] Alireza Makhzani, Jonathon Shlens, Navdeep Jaitly, and Ian Goodfellow, "Adversarial autoencoders," *Computer Science*, 2015.

[14] Olli Silvn, Matti Niskanen, and Hannu Kauppinen, "Wood inspection with non-supervised clustering.," *Machine Vision & Applications*, vol. 13, no. 5-6, pp. 275–285, 2003.

[15] C. E Shannon, "A mathematical theory of communication," *Bell Labs Technical Journal*, vol. 27, no. 4, pp. 379–423, 1948.

[16] Nobuyuki Ohtsu, "A threshold selection method from gray-level histograms," *IEEE Transactions on Systems Man & Cybernetics*, vol. 9, no. 1, pp. 62–66, 2007.

[17] M Donoser and H Bischof, "Using covariance matrices for unsupervised texture segmentation," in *International Conference on Pattern Recognition*, 2010, pp. 1–4.

[18] Jianming Zhang and Stan Sclaroff, "Saliency detection: A boolean map approach," in *IEEE International Conference on Computer Vision*, 2013, pp. 153–160.

[19] Maximiliano Montenegro, Jay I. Myung, and Mark A. Pitt, "Crackit - an image processing toolbox for crack detection and characterization," in *IEEE International Conference on Image Processing*, 2015, pp. 798–802.

[20] Martn Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S Corrado, Andy Davis, Jeffrey Dean, and Matthieu Devin, "Tensorflow: Large-scale machine learning on heterogeneous distributed systems," 2016.

[21] E Shelhamer, J. Long, and T Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.