International Journal of Mechanical and Production Engineering, ISSN(p): 2320-2092, ISSN(e): 2321-2071
Volume- 8, Issue-7, Jul.-2020, http://iraj.in

THE INDUSTRIAL APPLICATION AND THE METHOD OF IMPROVEMENT OF GANOMALY

[1]TA-WEI TANG, [2]WEI-HAN KUO, [3]JAUH-HSIANG LAN, [4]CHIEN-FANG DING, [5]HAKIEM HSU, [6]HONG-TSU YOUNG

[1,2,3]Department of Mechanical Engineering, National Taiwan University, Taipei 10617, Taiwan
[4]Taiwan Instrument Research Institute National Applied Research Laboratories Hsinchu, Taiwan
[5]3DFAMILY Technology Co., Ltd.
[6]Department of Mechanical Engineering, National Taiwan University, Taipei 10617, Taiwan
E-mail: [1]f06522717@ntu.edu.tw, [2]r07522716@ntu.edu.tw, [3]r07522749@ntu.edu.tw, [4]1909003@tiri.narl.org.tw, [5]hakiem@3dfamily.com, [6]hyoung@ntu.edu.tw.

Abstract - In recent years, deep learning technology, a well-developed image recognition method, has introduced to improve the quality of detection and adaptability to the diverse situation in the field of automatic optical inspection (AOI). However, the available data is highly imbalanced towards normality (i.e. a large amount of normal data and a small amount of abnormal data). It becomes a major challenge in the process of industrial application of deep learning; therefore, GANomaly, one of the best deep learning model of anomaly detection proposed in 2018, has drawn great attention from scientists. By learning the normal data, the generator of GANomaly can provide good quality fake images with a probability distribution which is similar to the input image. In addition, it has a discriminator to extract the image information. The residual score between the input image and the fake image created by the generator is a critical factor in anomaly detection. Nevertheless, the detection capabilities of GANomaly in industrial inspection have not been thoroughly discussed and verified. In this paper, the public industrial inspection dataset MVTec Anomaly Detection (MVTec AD) and wood samples from actual production lines were used to examine the performance of GANomaly. The situation of mixing abnormal data into the training data was fully investigated. Additionally, the area under the curves (AUCs) of sample inspection were improved by modifying different residual score to inspect industrial data.

Keywords - Deep Learning, Automatic Optical Inspection (AOI), GANomaly, Wood Inspection.

INTRODUCTION

Anomaly detection has become an important deep learning research topic in recent years. Since the convolutional neural network (CNN) was proposed, deep learning has achieved outstanding performance in image recognition [1–3], and has been successfully implemented in various inspection tasks [4–6]. However, in the industrial inspection, the issue of sample imbalance (i.e. the proportion of normal and abnormal samples in the total samples have significant difference) should be considered and carefully tackled. Beside, uncommon defects are very difficult to collect. In order to break through the bottleneck of deep learning in the application of industrial inspection, anomaly detection has become an important research topic in recent years.

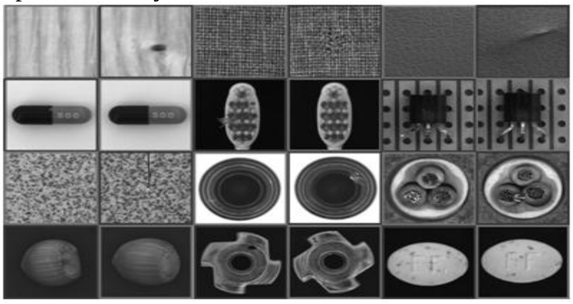

Fig. 1 Subsample of the dataset used in this study, including MVTec AD and Wood dataset on the production line. The gray frames and black frames are normal data and abnormal data, respectively.

Fortunately, GAN-based anomaly detection has outstanding performance in the anomaly detection dataset [7–9]. In comparison with traditional classifier-based deep learning models, GAN-based anomaly detection only requires to learn normal samples and is able to identify abnormal samples. GAN-based anomaly detection has been increasingly popular for the past few years due to the considerable difficulty in collection of defects in practical inspection. Scientists have successfully improved the performance of GAN-based anomaly detection deep learning models. Schlegl et al. [7] developed the architecture of Ano-GAN and utilized the characteristics of GAN for anomaly detection with the method of using the residual score as a defect detection method, which was superior to other methods. The result was also employed and enhanced by the subsequent GAN-based anomaly detection network [8, 9].

Akcay et al. [9] developed a convincing network for anomaly detection called GANomaly, and an encode-decode-encode architecture as a generator and an encoded architecture as a discriminator were used. Because of its great architecture, the network was successfully trained, and the performance of anomaly detection was better than Ano-GAN [10]. GANomaly has an impressive capability in the handwritten digit recognition dataset, object recognition dataset, and X-ray image recognition dataset. Furthermore, the processing speed is significantly improved compared with Ano-GAN. Nowadays, GANomaly is one of the best performing GAN-Based anomaly detection models.

However, it is necessary to verify the inspection capability of the model in industrial testing samples. For example, the residual score of GANomaly is the L1 distance of two vectors after the discriminator extracts the feature of the input image and the generated image, whereas it may not be the best residual score for all detection items. In this paper, the test detection capability of GANomaly with public industrial inspection dataset, improvement on the detection capability of GANomaly through different definitions of residual scores, and verification of the results with actual wood detection were investigated and discussed.

II. RELATED WORK

A. Generative Adversarial Networks (GAN)

Generative adversarial network is an unsupervised learning neural network [11]. This neural network learns to generate images with a similar probability distribution as the training data. It uses game theory to design the loss function of the neural network so that the network can compete with each other during training. Generative adversarial networks generally have a generator and a discriminator. The former generates the image data similar to the training data (i.e. fake image) through the structure of the neural network, while the latter generally includes the encode structure. It can identify the real image and the fake image. By training with the loss function and the optimizer, the training process forms with the concept of a zero-sum game. The generator and the discriminator compete with each other during the training process until the training reaches an equilibrium point. However, it is still challenging to reach a balance point in practical training. Therefore, several studies have tackled this issue using Wasserstein loss to improve the training results of GANs and stability of GAN training to avoid overfitting [12, 13].

B. Anomaly Detection

The problem of imbalance samples exists in many cases of inspection. The difficulty of abnormal sample collection makes anomaly detection technology an important development direction in deep learning. Paul et al. [14] proposed an anomaly detection dataset, i.e. MVTec AD, specifically used in industrial detection, and tested it with various anomaly detection methods [14]. The method included Generative Adversarial Networks (GAN), Deep Convolutional Auto encoders, features of pre-trained Convolutional Neural Networks (CNN) and traditional methods. However, the research showed that these methods still have much room for improvement in the identification of industrial inspection datasets.

C. GAN-based Anomaly Detection

GAN-based anomaly detection is an anomaly detection method which designed by utilizing the characteristics of GAN. Samet et al. [9] used GANomaly, which is one of the GAN- based anomaly with the best detection capabilities in recent years. The architecture is shown in Fig. 2. It consists of a generator, a discriminator, and an encoder. Through the training of the generator and the discriminator, the generator has a good ability to create the image which has similar probability distribution with input image, and the discriminator has a good feature extraction and recognition capability. By training the normal data, the generator can create the image which has similar distribution with normal image. Then, the original images and the fake images generated by the generator input to the discriminator for feature extraction. After feature extraction, the two images are encoded to two vectors. They defined the L1 distance of the two vectors after the feature extraction as the residual score. A higher residual score indicates that the sample is abnormal, while a lower residual score indicates that the sample is similar to the normal sample.

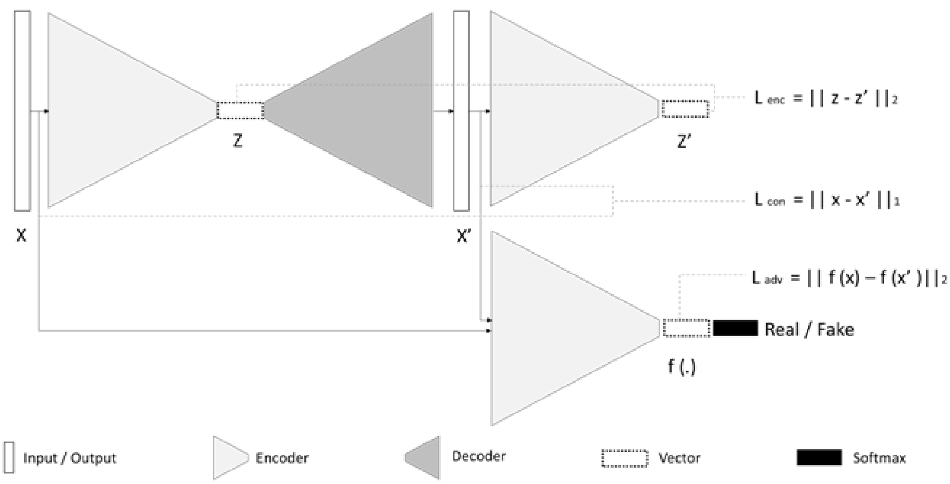

Fig. 2 Pipeline of the GANomaly for anomaly detection

In this study, GANomaly was trained and verified with the anomaly detection industrial dataset MVTec AD, and the abnormal samples were mixed into normal samples. The purpose of the experiment was to analyze the anti-noise capability of GANomaly under noisy labeled training data. In addition, in terms of different characteristic of samples (texture samples or objected samples), the best residual score definitions were searched, which consisted of L1 distance and L2 distance of the two images after feature extraction and the L1 distance between original image and fake image. Furthermore, the wood inspection samples on the actual production line was collected with a line scan camera. This was used to verify the feasibility of our proposed strategy for defining the best residual score and the detection ability of GANomaly on the actual production line.

III. METHOD

A. Background

As one of the best-performing GAN-based anomaly detection models, GANomaly is well-performed in handwritten digit recognition dataset MNIST, object recognition dataset CIFAR, and X-RAY security detection datasets UBA and FFOB. However, the networks may have different performance in different situations. The detection capability of GANomaly in industrial inspection has not been verified, and there may be more appropriate definitions of residual score for detection in certain tasks.

In addition, the defective samples are unintentionally mixed into training data which may cause a decrease in detection capability of the model. A series of experiments for the above problems were designed to verify the industrial detection capability of GANomaly in this study, including improving the performance on samples, and investigating the problems of mixing abnormal data into normal training data with images of actual production lines.

B. Datasets

In order to verify the performance of GANomaly in industrial inspection, two sets of datasets were used as follow:

(1) MVTec AD

The dataset collected by the MVTec Software GmbH team [14]. It contains 15 categories: Five of them are texture categories and ten are object categories. Each category has training data of normal samples and test data of normal samples and abnormal samples. It is an industrial dataset specially used for anomaly detection. There are 3,629 training images and 1,725 test images. The image resolutions are ranging from 700 x 700 to 1024 x 1024, and the image channels are single channel and three channels. 12 categories with three channels were used in this study.

(2) Wood dataset on the production line

The wood detection samples in the production line were scanned with a line scan camera. It contains six types of classes such as good products, chalk, holes, black, knots, and others. The subsamples are shown in the Fig. 3.

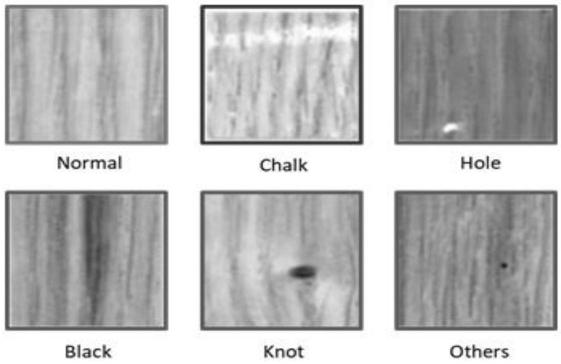

Fig. 3 Wood dataset on the production line. It contains normal and 5 defect classes.

There are 3,075 pieces of normal images for training and 740 pieces of test images, including good and bad samples. The resolution is 256 x 256, and all of the images are three channels.

C. Method

In order to systematically verify the issue regarding the performance of GANomaly in the industrial dataset, anti-noise ability, detection improvement of GANomaly under different definitions of residuals, and capability to inspect actual production line samples, the following three parts of the experiment were designed.

(1) Verification of the detection capability of GANomaly with industrial inspection datasets

To verify the performance of GANomaly in industrial inspection datasets, GANomaly was used to train and test 12 kinds of categories in MVTec AD, including four texture categories (carpet, leather, tile and wood) and eight object categories (bottle, cable, capsule, hazelnut, metal nut, pill, toothbrush and transistor). In addition, the abnormal samples were mixed with 0.03 times, 0.06 times, and 0.09 times of the original training data into the training data. These training data were used to verify the anti-noise ability of the model.

(2) Improvement of the detection capability of GANomaly through the different definition of residual scores

To make GANomaly have the best detection capability in each categories, the following three different definitions of residual score were used to detect the dataset:

(a) L1 distance with two vectors

After inputting the image P into the generator to generate a fake image Q, the original input P and the fake image Q were inputted to the discriminator to obtain two vectors $\bar{p}$ and $\bar{q}$. The L1 distance between the two vectors $\bar{p}$ and $\bar{q}$ was used as the residual score. The residual score is defined as

$$D_1\,(\bar{p}, \bar{q}) = \| \, \bar{p}, \vec{q} \, \|_1 = \sum_{i=1}^{n} |\, p_i - q_i \,| \qquad (1)$$

Where $\vec{p} = (p_1, p_2, \dots, p_n)$ and $\vec{q} = (q_1, q_2, \dots, q_n)$.

(b) L2 distance with two vectors

After inputting the image P into the generator to generate a fake image Q, the original input P and the fake image Q were inputted to the discriminator to obtain two vectors $\vec{p}$ and $\vec{q}$. The L2 distance between the two vectors $\vec{p}$ and $\vec{q}$ was used as the residual score. The residual score is defined as

$$D_2(\vec{p}, \vec{q}) = \sqrt{\sum_{i=1}^{n} (p_i - q_i)^2} \tag{2}$$

Where $\vec{p} = (p_1, p_2, \dots, p_n)$ and $\vec{q} = (q_1, q_2, \dots, q_n)$.

(c) L1 distance with two images

The image P was inputted the generator to generate a fake image Q. The L1 distance of the original input P and the false image Q was used as the residual scores, which is defined as

$$D_3(P, Q) = \sum_{(r,c)\in D} | P(r, c) - Q(r, c) | \tag{3}$$

Where r is the row index and c is the column index of the images P and Q.

(3) Verification of GANomaly with wood dataset on the production line

The wood dataset collected in the actual production line was used to verify the detection capability of GANomaly. The dataset consists of six classes, which were good, chalk, holes, black, knots, and others. After training, different definitions of residual scores were applied to detection and verified whether the results were consistent with the conclusions in the previous section.

D. Training Details

The GANomaly was used for training and detection in this paper. In order to make training fast and effective convergence, Adam was used as the optimizer. The learning rate was set to 0.001. In addition, to train GANomaly, the weights of each loss function were needed. The loss function is defined as:

$$L = \lambda_{adv} L_{adv} + \lambda_{con} L_{con} + \lambda_{lat} L_{lat} \tag{4}$$

The training parameters were set to $\lambda_{adv} = 20$, $\lambda_{con} = 1$, $\lambda_{lat} = 1$, and training steps were set to 20,000 for training. However, in most cases, the network learned sufficient information within less training steps. Therefore, the model with the best detection result was considered as the verification model. The experiment was carried out by using Intel i7-9700K 3.6 GHz CPU and Nvidia RTX 2080Ti 11GB GPU under the deep learning framework of Keras for training and verification.

E. Evaluation

The performance was evaluated by the area under the curve (AUC) of the receiver operating characteristics (ROC) in this study. The vertical axis of the ROC curve is true positive rate (TPR), and the horizontal axis is false positive rate (FPR). The curve is formed by connecting the TPR and FPR with different detecting threshold. When the AUC is closer to 1, it means that its detection capability is better. AUC is an effective way to evaluate the detection capability of the binary detection method, and it is also one of the commonly used methods for deep learning [8, 9, 15].

IV. RESULT

A. Verification of the detection capability of GANomaly with industrial inspection datasets

Table. I summarizes the AUC for training GANomaly without abnormal samples and with a small amount (<10%) of abnormal samples with different proportions. Without abnormal samples mixed into the training data, the AUCs were in a range of 0.671–0.92 for the MVTec AD dataset. There was the highest value (0.92) in Wood category, while the lowest value (0.671) in Pill category. It can be observed that this model had a large range in detection capability for different categories in industrial detection.

On the other hand, in the case where a small amount (<10%) of defective products were mixed into the training data, the impact on AUC was marginal. The reason is that the difference between the defective product image and the good product data was not very large, and the probability distribution of the entire image was still similar. Therefore, GANomaly had higher resistance to such noise in the training data. It also indicates that when GANomaly was actually applied to industrial inspection, a small amount of defective products is mixed accidentally into the good samples for training would not make the inspection system collapse.

	Proportion of abnormal samples in training data			
Categories	0	0.03	0.06	0.09
Bottle	0.794	0.775	0.861	0.773
Cable	0.711	0.694	0.705	0.675
Capsule	0.721	0.663	0.648	0.667
Carpet	0.821	0.581	0.879	0.837
Hazelnut	0.874	0.907	0.964	0.922
Leather	0.808	0.717	0.738	0.85
Metal nut	0.694	0.746	0.687	0.744
Pill	0.671	0.665	0.665	0.675
Tile	0.72	0.727	0.7	0.711
Toothbrush	0.7	0.784	0.622	0.723
Transistor	0.808	0.78	0.792	0.793
Wood	0.92	0.938	0.902	0.943

Table. I The AUCs of each category with different proportion of abnormal samples in training data.

B. Improvement of the detection capability of GANomaly through the different definition of residual scores

In this part, three distances were applied to define the residual score. MVTec AD dataset was used to detect these distances. Distance 1 represents L1 distance of the two vectors after feature extraction, as defined in equation (1). Distance 2 represents L2 distance of the two vectors after feature extraction, as defined in equation (2). Distance 3 represents L1 distance of the input image and the fake image, as defined in equation (3). Table. II presents the categories with higher AUC detected by using distance 3. It is clearly seen that in pill, bottle and hazelnut categories (object samples), the detection capability of distance 3 was significantly better than the distance 1. This was due to the fact that the generator was able to reconstruct the object samples much better than the texture samples with low variation in object samples. In the case of detecting object samples, feature extraction made data loss, resulting in a decrease in AUC.

On the other hand, Table. III presents the categories with higher AUC detected by using distance 1. It can be seen that in Carpet, Leather and Tile categories (texture samples), the detection capability of distance 1 was significantly better than distance 3. The reason is that when the sample was variable, the generator would only reconstruct the probability distribution of the sample image. It indicated that the generator cannot reconstruct the texture samples as well as object samples. In these case, the feature extractor can extract the significant feature, leading the capability of the network better. In terms of distance 2, since distance 2 was highly positively related to distance 1, the calculation of L2 made the smaller component value disappear. It made the capability of detection by using distance 2 was very close to that of using distance 1.

C. Verification of GANomaly with wood dataset

The actual production line wood dataset was used to train GANomaly. Table IV shows The AUC of wood dataset on the production line using different residual score.

Categories	Residual Score Distance 1	Distance 2	Distance 3
Pill	0.671	0.67	**0.778**
Bottle	0.794	0.800	**0.871**
Hazelnut	0.874	0.841	**0.948**
Transistor	0.808	0.807	**0.852**
Wood	0.920	0.922	**0.925**
Cable	**0.711**	0.695	**0.711**

Table. II The AUCs of each category of MVTec AD dataset with different residual score.

Note: This table only shows the categories which were detected better by using distance 3 than using distance 1 and distance 2. The highest AUCs of each row are shown in bold.

Categories	Residual Score Distance 1	Distance 2	Distance 3
Carpet	**0.821**	0.820	0.392
Leather	**0.808**	0.801	0.511
Tile	**0.720**	0.706	0.538
Metal nut	**0.694**	0.658	0.556
Capsule	**0.721**	0.698	0.662
Toothbrush	**0.700**	0.654	0.642
Cable	**0.711**	0.695	**0.711**

Table. III The AUCs of each category of MVTec AD dataset with different residual score.

Note: This table only shows the categories which were detected better by using distance 1 than using distance 2 and distance 3. The highest AUCs of each row are shown in bold.

It can be seen that GANomaly performed better by using distance 1 than distance 3.This was because the wood detection samples of the actual production line were texture samples which increased the detection ability in feature extraction. It is worth mentioning that there was also a wood dataset in MVTec AD, but the detection AUC by using distance 1 and distance 3 was very similar. The reason is the wood defects in MVTec AD mostly occupied a large area. Therefore, the network could tell abnormal samples from normal sample easily by using either distance 1 or distance 3 for detecting the sample. However, the defect area of wood dataset, which collected on the production line, was small. This was caused by the wood texture sample which was necessary to use the feature extractor to increase the capability of detection. In addition, an AUC of 0.915 indicated that the detection ability of wood samples in actual production lines was remarkable.

V. CONCLUSION

This paper has verified the capability of GANomaly in industrial inspection by using MVTec AD and wood inspection data from actual production lines. The result indicated that the detection performance of GANomaly in each sample was distinctly different. GANomaly had a great anti-noise ability for the situation where a small number of defective products were mixed into the training data. On the other hand, using distance 3 as the residual score to detect object samples was better than using the original residual score definition of GANomaly.

Categories	Residual Score Distance 1	Distance 2	Distance 3
Wood	**0.915**	0.914	0.849

Table IV. The AUC of wood dataset on the production line using different residual score.

Note: The highest AUC is shown in bold.

Besides, in the part verified GANomaly by using the wood samples of the actual production line, the AUC increased to 0.915 which indicated that the performance in this dataset was convincing.

REFERENCE

[1] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition", arXiv.org, 2014.

[2] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning", arXiv.org, 2016.

[3] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition", arXiv.org, 2016.

[4] X. Bian, S. Lim and N. Zhou, "Multiscale fully convolutional network with application to industrial inspection", 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), 2016.

[5] D. Weimer, B. Scholz-Reiter and M. Shpitalni, "Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection", CIRP Annals, vol. 65, no. 1, pp. 417-420, 2016.

[6] R. Ren, T. Hung and K. Tan, "A Generic Deep-Learning-Based Approach for Automated Surface Inspection", IEEE Transactions on Cybernetics, vol. 48, no. 3, pp. 929-940, 2018.

[7] T. Schlegl, P. Seeböck, S. Waldstein, U. Schmidt-Erfurth and G. Langs, "Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery", Lecture Notes in Computer Science, pp. 146-157, 2017.

[8] H. Zenati, C. S. Foo, B. Lecouat, G. Manek, and V. R. Chandrasekhar, “Efficient gan-based anomaly detection,” arXiv preprint arXiv:1802.06222, 2018.

[9] S. Akcay, A. Atapour-Abarghouei and T. Breckon, "GANomaly: Semi-supervised Anomaly Detection via Adversarial Training", Computer Vision – ACCV 2018, pp. 622-637, 2018.

[10] F. Di Mattia, P. Galeone, M. De Simoni and E. Ghelfi, "A Survey on GANs for Anomaly Detection", arXiv.org, 2019.

[11] B. Kiran, D. Thomas and R. Parakkal, "An Overview of Deep Learning Based Methods for Unsupervised and Semi-Supervised Anomaly Detection in Videos", Journal of Imaging, vol. 4, no. 2, p. 36, 2018.

[12] S. Ioffe and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", arXiv.org, 2015.

[13] M. Arjovsky, S. Chintala, and L. Bottou, “Wasserstein generative adversarial networks,” in Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 06–11 Aug 2017, pp. 214–223, Aug 2017.

[14] P. Bergmann, M. Fauser, D. Sattlegger and C. Steger, "MVTec AD -- A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9592-9600, 2019.

[15] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, “Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 10265 LNCS, pp. 146–147, 2017.

★★★