IEEE Transactions on Automation Science and Engineering (T-ASE) paper, presented at ICRA 2024, Yokohama, Japan.
Cite as T-ASE paper.

Multi-category Decomposition Editing Network for the Accurate Visual Inspection of Texture Defects

He Zhu[1], Junyi Li[1], Hua Yang[1], Jiankui Chen[1] and Zhouping Yin[1]

***Abstract*— Spotting blemished areas automatically on a textured surface is a particular challenge, as both nominal and defective surface samples are inconsistent in large-scale industrial manufacturing. The most efficient solution uses the memory bank extracted from the nominal samples to detect outliers. We approach our strategy, the multi-category decomposition editing network (MCDEN), from a similar viewpoint. Notably, we do not use defect-free samples. Instead, we use virtual results to construct a defect library. MCDEN decomposes abnormalities to basic elements from the library while editing outlier features to reconstruct the texture normality, offering a rational segmentation map through decomposition and reconstruction. Based on the strategy, MCDEN is more interpretable than most neural network methods since interpretability is particularly important in industry to ensure stability. Experiments on texture surface samples from the MVTAD dataset confirm the efficacy of MCDEN with a pixel-level AUC score of 96.6%. In other experiments collected from semi-manufactured inkjet printing OLED panels, MCDEN demonstrates competitive results with a 99.2% detection rate and rapid real-time detection capability.**

I. INTRODUCTION

For high reliability inspection, a growing number of industrial visual anomaly detection methods have relied on fitting a model with a nominal distribution using autoencoding networks, generative adversarial networks (GANs), or other unsupervised methods with only nominal samples. A popular strategy that emerged recently is feature matching between the test and normal samples which is referred to as an embedding-based method. It only stores features extracted from normal samples and forms a memory bank, introducing an standard to judge the distance between the images to be detected and the memory. To extract better features, it is usually necessary to use complex pre-trained models. However, the detection speed when running a network is time-consuming due to the complexity of the feature extraction and complicated feature matching.

Given that the existing methods are advanced yet lacking, we propose a segmentation-based method, the multi-category decomposition editing network (MCDEN). Unlike a memory bank using normal samples, MCDEN uses a defect library with abnormal data to recognize and decompose defect regions into assumed fundamental conversions, as shown in Fig. 1. Meanwhile, to increase the utilization rate of defect decomposition information, a transformation model, such as the inverse transformation model, is designed to address defect region faults by editing features according to basic conversions. MCDEN is based on the assumption that conversion element combination defects, such as specific angle rotations and shape deficiencies or translucent masks with an alpha channel. This assumption comes from observations of industrial production, in which the merchandise surface is classified by texture and defects are small and rare.

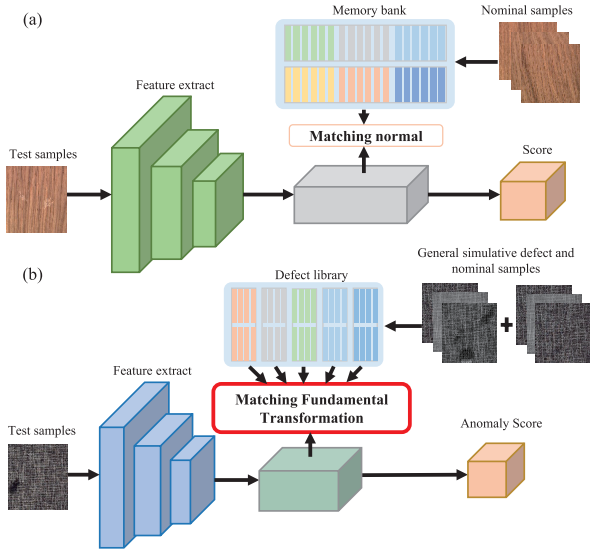

Fig. 1. Comparison between conventional memory bank and proposed defect library: (a) operation of memory bank, (b) operation of defect library.

Experiments on data collected from the open-source MVTAD dataset demonstrate the effectiveness of MCDEN with a pixel-level AUC score of 96.6%. Furthermore, a dataset collected from the working production line of organic light-emitting diode (OLED) luminescent layer ink printing is applied, and a 99.2% score for the detection rate is achieved. The cost is only 0.097 s for one pattern, demonstrating that MCDEN is satisfactory for real-time defect detection in production lines.

To summarize, the main contributions of this paper are as follows:

- The MCDEN defect detection method is proposed to solve the textured surface product real-time inspection problem. This method achieves competitive performance on the MVTAD dataset and semi-finished OLED dataset. In addition, MCDEN is more interpretable than most other methods, making it more appropriate in industrial manufacturing.

*This work was supported by the Joint Funds of National Natural Science Foundation of China U22A20208, the Key Research and Development Program of Guangdong Province under Grant 2022B0202010001-2, the project of Foshan Science and Technology Bureau under Grant 2020001006509.

[1]The authors are with the State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science and Technology, Wuhan 430074, China.

- A strategy that extended on memory bank, from another perspective, decomposes texture defects into fundamental conversion set by the defect library proposed.
- A negative sample generation method that is similar to CutMix allows normal and abnormal areas to be distinguished, achieving the purpose of improving the accuracy of defect decomposition.

II. Related Works

In industrial production, because only a few defect samples can be used for model training, unsupervised inspection methods, including methods based on background reconstruction, and embedding, popular in industrial fields.

Background reconstruction-based methods aim at reconstructing images to be detected into defect-free background images through networks using autoencoders (AEs) [1] [2] [3] and generative adversarial networks (GANs) [4] [5] [6] [7] and detecting anomaly areas by calculating the residuals. In the early stages, MS-FCAE [8] was proposed to explore texture defects in multiple dimensions by reconstructing a multi-scale background. However, due to the great power of the above mentioned networks [9] [10], they recur well not only in background regions but also possibly in defect regions. To reduce the influence, RIAD [11], which attempts to cover abnormal regions using masks obtained by an operation generating multi-scale complementary masks. However, the models still encounter errors because the nature of defects is not understood. MCDEN extends to existing reconstruction models and starts from the defect elements.

Embedding-based methods [12] [13] extract abstraction information of nominal samples at a high level, inferring by comparing the distance between the features of test images and nominal images. An anomaly score will be generated to indicate which regions appear to be defects, and a pixel-level result can be output. These methods always rely on a network pretrained on ImageNet to extract sufficient features and perform complex feature matching, which costs a large amount of time. Memory-guided normality for anomaly detection (MNAD) [14] is based on using a memory model to detect anomalies by updating the memory bank during inference. In a new framework for patch distribution modeling (PaDiM) [15], patch features of the input images are extracted, and abnormalities are detected by nominal sample fitting of the probability representation using a multivariate Gaussian distribution. However, the complexity of these methods makes them time-consuming. Efficiency is critical in industrial production, MCDEN has taken relevant ideas and achieved much higher efficiency than such methods while ensuring accuracy.

Overall, MCDEN draws on the experience of its predecessors and optimizes a strategy for industrial application.

III. METHOD

A. MCDEN Architecture

The whole network construction is as shown in Fig. 2. The defect library projection network is referred to as the multi-category library module (MCLM), the background image reconstruction network is referred to as the multi-scale convolution edit module (MSCEM), and the simulated defect generation network is referred to as the random composition generation module (RCGM).

A defect detection network was developed with the idea of decomposing, importing or projecting images in the defect library trained in advance. Then, we use the set of basic conversion scores to define or learn the corresponding inverse transformation to reconstruct the background, further obtaining the residual between the original data and reconstructed background data. Combined with the defect score set and residual, will be exported and reflect the final product of the whole inspection.

B. Multi-category Library Module

The MCLM aims to transition defects through basic conversion , approximately locate the area anomaly, and operate on the latent feature Z^a through the encoder:

$$Z^a = f_e\left(I_a; \theta_e\right) \tag{1}$$

where $Z^a \in R^{W_l \times H_l \times C_l}$; W_l, H_l, and C_l denote the width, height, and channel of the latent feature, respectively. $f_e\left(\cdot\right)$ denote the encoder function.

Decomposition begins following the acquisition of Z^a, The MCLM contains K preset defect libraries for the hypothetical transformation: $l_i, i = 1, \ldots, K$. Each library can be taken as a matrix $l_i \in R^{M \times C_l}$, where M denotes the capacity of each library. Thus, there are M vector entries in a library, and each entry $l_{i,j} \in R^{1 \times C_l}, j = 1, \ldots, M$ is the same length as the channel of the latent feature.

To measure the distance between feature z^a and library vector $l_{i,j}$ stably and simply, cosine similarity is used for the index, where z^a is a single feature in the Z^a, and the maximum similarity between z^a and all entries are employed as the similarity of this library d_i, where $\|\cdot\|_2$ denotes the Euclidean norm.

$$d_{i,j} = \frac{z^a {l_{i,j}}^T}{\|z^a\|_2 \|l_{i,j}\|_2} \tag{2}$$

$$d_i = \max_j d_{i,j} \tag{3}$$

The similarity between z^a and each library l_i can be taken as the anomaly score of the corresponding basic transformation, which can be regarded as completing the transformation, and score set $d = [d_1, \ldots, d_K]$ is obtained.

To improve the decomposition quality and reduce the coupling degree between transforms and increase the specificity of each basic transformation to improve generalization ability, an orthogonal constraint is applied.

$$L_o = \frac{1}{M^2 \binom{K}{2}} \sum_{1 \le u < v \le K} \left(P \left(L_u {L_v}^T\right) P^T\right) \tag{4}$$

where $\binom{\cdot}{\cdot}$ is a combinatorial number and $P \in R^{1 \times M}$ is an auxiliary row matrix that is used to compute the sum of all elements of the similarity matrix $L_u L_v$.

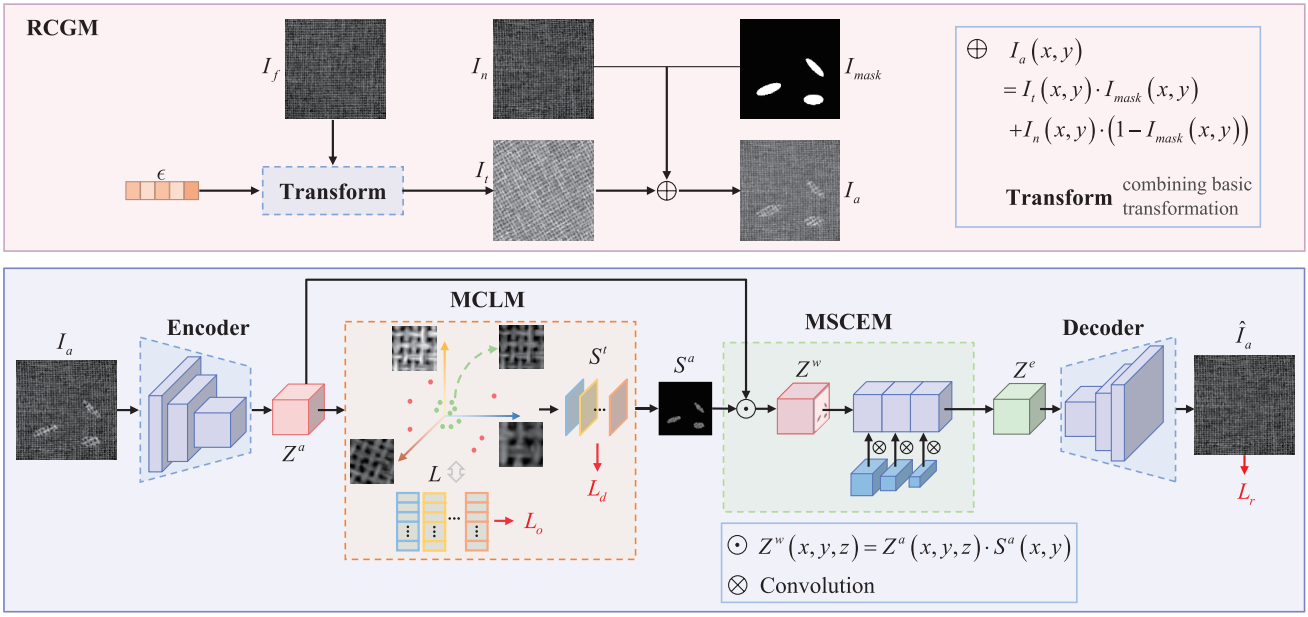

Fig. 2. The network architecture of the proposed MCDEN model. It consists of three components: the RCGM, the MCLM and the MSCEM.

Furthermore, the Euclidean norm is used for the training target of decomposition label $V^t \in R^{W_l \times H_l \times K}$ and transformation score maps S^t to standardize features for better handling based on transformation independence. Meanwhile, V^t can be obtained by downsampling the map V^{tu} from size $W \times H \times K$ to $W_l \times H_l \times K$, and V^{tu} obtained from the dimension extension calculation between Gaussian defect mask I_{mask} and random code ϵ, marked $f_{dex}(\cdot)$. In Formula (6), $\mathbb{E}[\cdot]$ denotes the expectation operation.

$$L_d = \mathbb{E}_{S^t}\left[\left\|V^t - S^t\right\|_2\right] \tag{5}$$

$$V^{tu}(x,y,k) = f_{dex}\left(I_{mask}(x,y), \epsilon(k)\right) \tag{6}$$

A final anomaly score map S^a is fused and normalized by Formula (7), which is based on orthogonal constraint L_o and decomposition constraint L_a, where $S^a \in R^{W_l \times H_l \times 1}$, $x_l = 1, \ldots, W_l$, and $y_l = 1, \ldots, H_l$. By computing the scores between each feature vector in Z^a and the basic transformations, the latent feature Z^a can be decomposed into basic transformation score maps S_i^t, where $S_i^t \in R^{W_l \times H_l \times 1}$ corresponds to each basic transformation. S^t denotes the whole transformation score map for the decomposition result, where $S^t \in R^{W_l \times H_l \times K}$, and S^a indicates the fused anomaly score and the defective region of the combined anomaly.

$$S^a(x_l, y_l) = \frac{\sum_{i=1}^{K} S_i^t(x_l, y_l)}{\max\left(\sum_{i=1}^{K} S_i^t(x_l, y_l)\right)} \tag{7}$$

In conclusion, the whole decomposition procedure of the MCLM can be condensed, as shown in Formula (8), where $f_l(\cdot)$ and θ_l represent the MCLM function and main parameters, respectively.

$$S^a = f_l(Z^a; \theta_l) \tag{8}$$

C. Multi-scale Convolution Edit Module

The MSCEM is a model that receives information that has been adjusted from the MCLM, exports defect-free background reconstruction data, and restrains or inversely transforms defect areas through the defect score feature to nominal feature editing.

An operation, the pixel-level product, aimed at locating abnormal features, as in Formula (9), is used. $Z^w \in R^{W_l \times H_l \times C_l}$, $z_l = 1, \ldots, C_l$ is the set of abnormal features and x_l, y_l, z_l are the components of W_l, H_l, C_l.

$$Z^w(x_l, y_l, z_l) = Z^a(x_l, y_l, z_l) \cdot S^a(x_l, y_l) \tag{9}$$

Due to the various sizes and shapes of the defect regions, convolution layers with different kernel sizes are employed to fill these areas by utilizing the information of the neighborhood texture background, as shown in Formula (10), where $Z^e \in R^{W_l \times H_l \times C_l}$, $f_c(\cdot)$ and θ_c denote the functions and parameters of the multi-scale convolution layers. As a result, defect-free feature Z^e is obtained from Z^a and S^a, and a texture background image $\hat{I}_a \in R^{W \times H \times 1}$ can be reconstructed by passing the Z^e to the decoder, as shown in Formula (11), where $f_d(\cdot)$ and θ_d denote the function of the decoder and the main parameters.

$$Z^e = f_c(Z^w; \theta_c) \tag{10}$$

$$\hat{I}_a = f_d(Z^e; \theta_d) \tag{11}$$

Constraints are added to ensure the reconstruction quality based on the Euclidean norm, as shown in Formula (12),

where I_n is the original defect-free image, $\hat{I}_a$ is the reconstructed texture background image.

$$L_r = \mathbb{E}_{I_n}\left[\left\|I_n - \hat{I}_a\right\|_2\right] \tag{12}$$

D. Random Composition Generation Module

The RCGM is designed to provide general defect samples for the whole network training. Mask generation involves locating a place randomly to generate defects using two-dimensional Gaussian weight maps I_{mask}.

$$\Phi^\iota(x,y) = A^\iota e^{-\left(a^\iota(\Delta x^\iota)^2 + b^\iota \Delta x^\iota \Delta y^\iota + c^\iota(\Delta y^\iota)^2\right)} \tag{13}$$

$$\begin{aligned} a^\iota &= \frac{\cos^2\theta^\iota}{2{\sigma_x^\iota}^2} + \frac{\sin^2\theta^\iota}{2{\sigma_y^\iota}^2}, \\ b^\iota &= \frac{\sin 2\theta^\iota}{2{\sigma_x^\iota}^2} - \frac{\sin 2\theta^\iota}{2{\sigma_y^\iota}^2}, \\ c^\iota &= \frac{\sin^2\theta^\iota}{2{\sigma_x^\iota}^2} + \frac{\cos^2\theta^\iota}{2{\sigma_y^\iota}^2} \end{aligned} \tag{14}$$

$$\Delta x^\iota = x - x_0^\iota, \Delta y^\iota = y - y_0^\iota \tag{15}$$

where $A^\iota, \theta^\iota, \sigma_x^\iota, \sigma_y^\iota, x_0^\iota, y_0^\iota$ where $\iota = 1, \ldots, N_a$, are Gaussian function parameters and represent the strength, rotation angle, dispersion of x, dispersion of y, center coordinate of x, center coordinate of y. N_a is the number of defect regions in one image, randomly limited in $1 \le N_a \le 3$. $\Phi^\iota \in R^{W \times H}$. x, y are the components of W and H. We limit these value, $0.3 \le A^\iota \le 1$, $-90^\circ \le \theta^\iota \le 90^\circ$, $10 \le \sigma_x^\iota$, $\sigma_y^\iota \le 60$, $20 \le x_0^\iota \le W - 10$, $10 \le y_0^\iota \le H - 10$, and select them randomly. For each naturally generated region I_m^ι, a random threshold truncation is applied on map Φ^ι, as in Formula (16), where $I_m^\iota \in R^{W \times H \times 1}$, and T_m is random in $[0.3, 0.6]$. Then, the Gaussian defect mask I_{mask} is generated from I_m^ι, as in Formula (17).

$$I_m^\iota(x,y) = \begin{cases} 1, \Phi^\iota(x,y) > T_m \cdot \max \Phi^\iota \\ 0, otherwise \end{cases} \tag{16}$$

$$I_{mask}(x,y) = \max_\iota\left(I_m^\iota(x,y)\right) \tag{17}$$

The core of the RCGM is the content generation, which affects the defect libraries in the MCLM, and the content randomly originates from the reference texture image I_f. For the observation of a production line, the defects can be summarized as a difference in brightness or structure from the texture background, including brightening, darkening, enlarging, narrowing and rotating operations. The transformations above 5 are related to random code $\epsilon = [\epsilon_1, \ldots, \epsilon_5]$, and each element of controls the degree of one basic transformation. In particular, to prevent the codes from canceling each other, one of the converse transformation codes $\epsilon_u, \epsilon_v, 1 \le u, v \le 5, u \ne v$ is set to 0 randomly.

Given random code ϵ, the defect content can be transformed from the reference images I_f. Transformations are performed sequentially with 3 factors: grayscale λ_g, scale λ_s, and rotation λ_r.

$$\lambda_g = \begin{cases} \alpha^b + c_1 \cdot \beta^b + \varepsilon^b, c_1 \ne 0 \\ \alpha^d + c_2 \cdot \beta^d + \varepsilon^d, c_2 \ne 0 \\ 1, otherwise \end{cases} \tag{18}$$

$$\lambda_s = \begin{cases} \alpha^e + c_3 \cdot \beta^e + \varepsilon^e, c_3 \ne 0 \\ \alpha^n - c_4 \cdot \beta^n - \varepsilon^b, c_4 \ne 0 \\ 1, otherwise \end{cases} \tag{19}$$

$$\lambda_r = \begin{cases} \alpha^r + c_5 \cdot \beta^r + \varepsilon^r, c_5 \ne 0 \\ 0^\circ, otherwise \end{cases} \tag{20}$$

where $\alpha^b = \alpha^e = 1.1, \alpha^d = \alpha^n = 0.9, \alpha^r = 10^\circ$ denotes the transformation minimum, $\beta^b = \beta^d = \beta^e = \beta^n = 0.5, \beta^r = 50^\circ$ denotes the transformation interval, $0 \le \varepsilon^b, \varepsilon^d, \varepsilon^e, \varepsilon^n \le 0.1, 0 \le \varepsilon^r \le 10^\circ$ represents the disturbance added in the transformations to improve sample reliability.

The transformation image I_t can be obtained by operating I_f through grayscale, scale, and rotation transformations, where $T(\cdot)$ is the content transformation function.

$$I_t = T(I_r; \lambda_g, \lambda_s, \lambda_r) \tag{21}$$

The defect sample I_a can be composed with the CutMix method using I_n, I_t and I_{mask}, as shown in Formula (22), where I_a, I_n, I_t, and $I_{mask} \in R^{W_l \times H_l \times 1}$, and I_n denotes the normal texture image.

$$\begin{aligned} I_a(x,y) = {} & I_t(x,y) \cdot I_{mask}(x,y) \\ & + I_n(x,y) \cdot (1 - I_{mask}(x,y)) \end{aligned} \tag{22}$$

E. Training and Testing Procedure

MCDEN training consists of three procedures: encoder and decoder pretraining, defect library training, and whole network learning.

To improve the representative feature extraction ability, encoder and decoder pretraining is needed on normal images I_n based on the constraint in Formula (23), where $Z^n \in R^{W_l \times H_l \times C_l}$ is the normal feature and $\hat{I}_n$ denotes the reconstructed image of I_n. The Adam optimizer is used with a learning rate of 0.001.

$$L_{pre} = \mathbb{E}_{I_n}\left[\left\|I_n - \hat{I}_n\right\|_2\right] \tag{23}$$

$$Z^n = f_e(I_n; \theta_e) \tag{24}$$

$$\hat{I}_n = f_d(Z^n; \theta_d) \tag{25}$$

Whole network training begins when the RCGM has generated the samples, defect library initialization has finished and convergence has occurred in encoder and decoder pretraining under constraint L, where λ_1, λ_2 and λ_3 are set to 10, 0.001, 0.1 respectively.

$$L = \lambda_1 L_r + \lambda_2 L_o + \lambda_3 L_d \tag{26}$$

In inferring time, the input images I_d will be reconstructed to generate the background image $\hat{I}_d$ and a rough prediction

TABLE I

THE AUC ON THE FIVE TEXTURES OF MVTAD [16]. BOLD AND UNDERLINED SCORES INDICATE THE BEST AND SECOND-BEST SCORES, RESPECTIVELY.

Model	Carpet	Grid	Leather	Tile	Wood	Mean
AnoGAN [4]	0.54	0.58	0.64	0.50	0.62	0.576
AE-L2 [1]	0.59	0.90	0.75	0.51	0.73	0.696
AE-SSIM [2]	0.87	0.94	0.78	0.59	0.73	0.782
OCGAN [17]	0.792	0.876	0.865	0.748	0.810	0.818
SPADE [12]	0.975	0.937	0.976	0.874	0.885	0.929
FCDD [18]	0.96	0.91	0.98	0.91	0.88	0.928
Patch SVDD [19]	0.926	0.962	0.974	0.914	0.908	0.937
CutPaste [20]	0.983	0.976	0.995	0.907	0.956	0.963
SBR [21]	0.95	**0.98**	0.98	0.95	0.93	0.96
PaDiM [15]	**0.991**	0.973	0.992	0.941	0.949	**0.969**
MCDEN	0.975	0.966	0.993	**0.978**	0.917	0.966

I_s will be obtained from I_d by upsampling the interim value anomaly score S^d. The final detection is fused by Formula (27), where $I_r \in R^{W_l \times H_l}$, $0 \leq \lambda_f \leq 1$, 0.7 is recommended, and $|\cdot|$ means length, or absolution.

$$I_r = \lambda_f \left| I_d - \hat{I}_d \right| + (1 - \lambda_f) I_s \tag{27}$$

For ease of use in industrial production, converting the defect probability to a binary result is optional.

IV. EXPERIMENTS

To verify the effectiveness of the MCDEN, experiments on texture categories collected from the MVTAD dataset have been conducted. Moreover, applying MCDEN to the ink printing dataset proves its capacity to be applied to industrial manufacturing.

A. Datasets

The main experiments on texture samples, including carpet, grid, leather, tile, and wood, are tested on the MVTAD dataset, a mainstream anomaly detection dataset focusing on real-world applications. Each category contains over 200 defect-free images for training and approximately 70 images with defects for inferring. To reduce computation operands, images were all converted to grayscale and resized to a resolution of 256 $\times$ 256 using MCDEN.

B. Comparative Experiments

To verify the performance of the proposed method, quantitative comparative experiments are conducted on the textures of the MVTec anomaly detection (MVTAD) dataset. The MCDEN is compared with several frequently used deep learning defect inspection methods, including AnoGAN [4], AE-L2 [1], AE-SSIM [2], OCGAN [17], SPADE [12], FCDD [18], Patch SVDD [19], CutPaste (3-way) [20], SBR [21] and PaDiM [15]. The quantitative results are shown in Table I. From the comparative experimental results, the proposed method obtains an average pixel-level AUC of 0.966, outperforming most of the compared methods. The overall score of MCDEN is second only to that of PaDiM with a small difference. Moreover, MCDEN obtains the best inspection performance among all the compared methods on the tile texture images. Some of the reconstruction and detection results of the proposed method on MVTAD are shown in Fig. 3.

TABLE II

ABLATION ANALYSIS FOR MCDEN ON THE CARPET TEXTURE.

Module	A	B	C	D	E
MCLM	✓	✓	✓	✓	
MSCEM	✓		✓		✓
Fusion	✓	✓			
AUC	0.975	0.964	0.960	0.944	0.941

C. Ablation Study Analysis

To analyze the efficacy of each MCDEN module, ablative experiments are conducted on the carpet category in the MVTec dataset, and the results are shown in Table. II. Each of the 5 experiments is assigned a letter, such as A, B, C, D, and E, and there are 3 kinds of assembly operations, including MCLM, MSCEM, and Fusion (score fusion). The effectiveness of the MCLM can be demonstrated by comparing column C with E in Table. II. Some of the decomposition results are shown in Fig. 5. For the results on the carpet image, the defect is decomposed into five basic categories. The scores of the darkening, narrowing, and rotation transformations are more obvious, which shows the basic components of this defect.

D. Basic Transformation Analysis

To verify whether the five transformations are necessary and determine whether there is a better combination to make the model lighter, experiments on the MVTAD dataset are conducted. These experiments include four combinations of the five transformations in Fig. 6 and three combinations with two opposing operations separately removed. Different texture surfaces have different requirements for basic transformations. According to the results, the accuracy is the highest when the five transformations are utilized.

E. Inference Speed

The MCDEN method is designed to detect defects in a surface approximated as a texture, achieving high accuracy and similar real-time performance. Experiments are carried out on a hardware platform with an NVIDIA RTX 2080 TI GPU. The time consumption when inferring SPADE, PaDiM, and MCDEN is 0.652 s, 0.635 s, and 0.097 s, respectively. When processing an image with a resolution of 256$\times$256, the processing time is six times faster, meaning that the MCDEN method is more appropriate for real-time industry production.

F. Application on Inkjet Printing

Inkjet printing equipment, as shown in Fig. 7, is developed to manufacture and inspect inkjet-printed OLEDs. The vision system in the equipment is mainly used to inspect defects on the substrate after inkjet printing caused by ink leakage and consists of two industrial color complementary metal oxide semiconductor (CMOS) cameras with 10 million pixel resolution and a resolving accuracy of 3.45 μm.

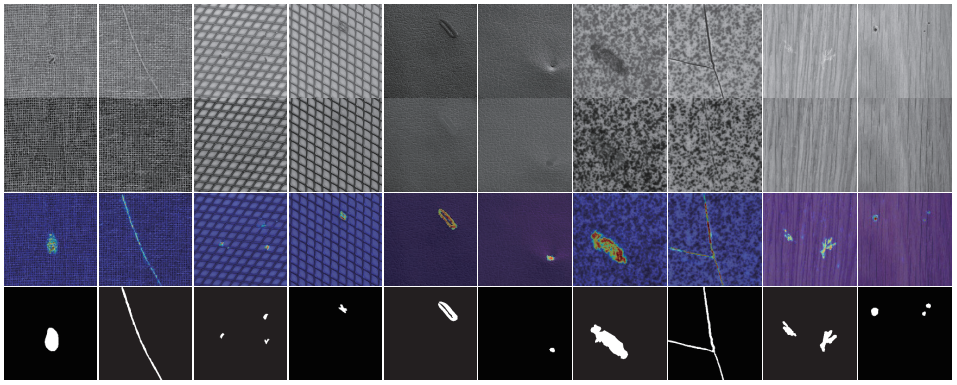

Fig. 3. The reconstruction and detection results of MCDEN on textures from MVTAD dataset. The rows represent the gray defective images, the reconstructed images, the detection results and the ground truth from top to bottom.

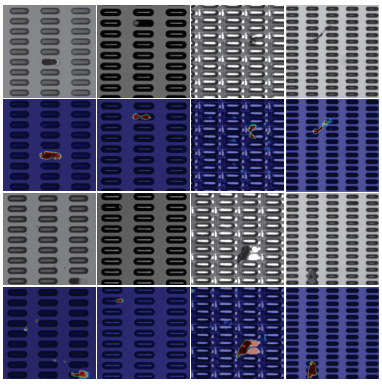

Fig. 4. Results of inkjet printing detection. Cols are small in volume, large ink volume, tight layout, low magnification classes.

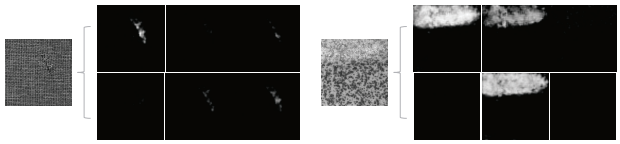

Fig. 5. The decomposition results on carpet and tile samples are shown in three columns to the right, and left two are the origin, each of them two groups are fused anomaly score and transition of brighten, darken, enlarge, narrow, rotate, sequenced from left to right and up to down: (a) a decomposition result of carpet, (b) a decomposition result of tile.

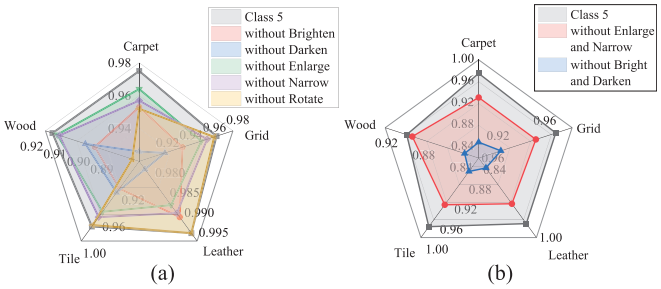

Fig. 6. Impact of different decomposition transformation combination: (a) results after removing a certain transformation, (b) results after removing a pair of opposite transformations

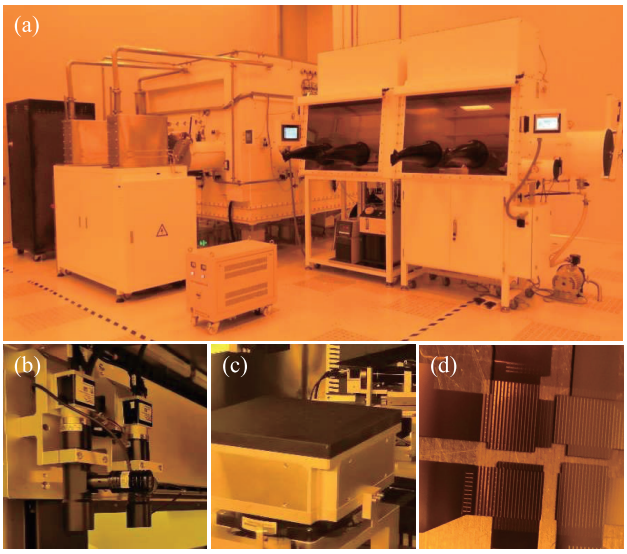

Fig. 7. Inkjet printing condition: (a) complete equipment, (b) vision system, (c) adsorption platform, (d) OLED illuminated substrate.

TABLE III
DETECTION RESULT OF MCDEN ON INKJET PRINTING DEFECTS.

State	Small ink volume	Large ink volume	Tight layout	Low magnification
Sample number	781	572	168	594
Correct Number	776	567	163	592
Detection Rate (%)	99.4	99.1	97.0	99.7

Experimental data collected from the vision system is applied to the MCDEN to satisfy the ink printing equipment requirements. There are 1000 defect-free inkjet printing images in the training set and 2115 defective images in the testing set. Among the defective images, there are 4 types of images. All images are cropped from 10 million images into patches with a resolution of 256×256, and a type 200×200 OLED substrate requires approximately 180 to 550 images for different layouts. Some detection results are in Fig. 4. As shown in Table III, MCDEN achieves a detection rate of 99.2% and a time cost of approximately 0.097 s for one patch and only 17.46 s to 53.35 s for the whole image.

V. CONCLUSIONS

MCDEN establishes a defect library to decompose anomalies to fundamental transformations and edits features extracted by an encoder to substitute or fill anomaly areas with the nominal texture to reconstruct a background image without defects. The MCDEN approach demonstrates that the decomposition of defects is viable by experiments on the MVTAD dataset, and 96.6% AUC is achieved. In addition, MCDEN has been applied to inkjet printing images. MCDEN requires less time, with a cost of 0.097 s for one patch and a detection rate of 99.2%, which satisfies the requirements of industrial manufacturing.

REFERENCES

[1] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *science*, vol. 313, no. 5786, pp. 504–507, 2006.

[2] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, "Improving unsupervised defect segmentation by applying structural similarity to autoencoders," in *VISIGRAPP*, 2019, pp. 372–380.

[3] D. T. Nguyen, Z. Lou, M. Klar, and T. Brox, "Anomaly detection with multiple-hypotheses predictions," in *International Conference on Machine Learning*. PMLR, 2019, pp. 4800–4809.

[4] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *International conference on information processing in medical imaging*. Springer, 2017, pp. 146–157.

[5] M. Sakurada and T. Yairi, "Anomaly detection using autoencoders with nonlinear dimensionality reduction," in *Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis*, 2014, pp. 4–11.

[6] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon, "Ganomaly: Semi-supervised anomaly detection via adversarial training," in *Asian conference on computer vision*. Springer, 2018, pp. 622–637.

[7] S. Pidhorskyi, R. Almohsen, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," vol. 31, 2018.

[8] H. Yang, Y. Chen, K. Song, and Z. Yin, "Multiscale feature-clustering-based fully convolutional autoencoder for fast accurate visual inspection of texture surface defects," *IEEE Transactions on Automation Science and Engineering*, vol. 16, no. 3, pp. 1450–1467, 2019.

[9] B. J. Wheeler and H. A. Karimi, "A semantically driven self-supervised algorithm for detecting anomalies in image sets," vol. 213. Elsevier, 2021, p. 103279.

[10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," vol. 40, no. 4. IEEE, 2017, pp. 834–848.

[11] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," vol. 112. Elsevier, 2021, p. 107706.

[12] N. Cohen and Y. Hoshen, "Sub-image anomaly detection with deep pyramid correspondences," *arXiv preprint arXiv:2005.02357*, 2020.

[13] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14 318–14 328.

[14] H. Park, J. Noh, and B. Ham, "Learning memory-guided normality for anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 14 372–14 381.

[15] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: a patch distribution modeling framework for anomaly detection and localization," in *International Conference on Pattern Recognition*. Springer, 2021, pp. 475–489.

[16] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 9592–9600.

[17] P. Perera, R. Nallapati, and B. Xiang, "Ocgan: One-class novelty detection using gans with constrained latent representations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 2898–2906.

[18] P. Liznerski, L. Ruff, R. A. Vandermeulen, B. J. Franks, M. Kloft, and K. R. Muller, "Explainable deep one-class classification," in *International Conference on Learning Representations*, 2020.

[19] J. Yi and S. Yoon, "Patch svdd: Patch-level svdd for anomaly detection and segmentation," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[20] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, "Cutpaste: Self-supervised learning for anomaly detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 9664–9674.

[21] C. Lv, F. Shen, Z. Zhang, D. Xu, and Y. He, "A novel pixel-wise defect inspection method based on stable background reconstruction," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, 2020.