Abstract—Reconstruction method based on the memory module for visual anomaly detection attempts to narrow the reconstruction error for normal samples while enlarging it for anomalous samples. Unfortunately, the existing memory module is not fully applicable to the anomaly detection task, and the reconstruction error of the anomaly samples remains small. Towards this end, this work proposes a new unsupervised visual anomaly detection method to jointly learn effective normal features and eliminate unfavorable reconstruction errors. Specifically, a novel Partition Memory Bank (PMB) module is proposed to effectively learn and store detailed features with semantic integrity of normal samples. It develops a new partition mechanism and a unique query generation method to preserve the context information and then improves the learning ability of the memory module. The proposed PMB and the skip connection are creatively explored to make the reconstruction of abnormal samples worse. To obtain more precise anomaly localization results and solve the problem of cumulative reconstruction error, a novel Histogram Error Estimation module is proposed to adaptively eliminate the unfavorable errors by the histogram of the difference image. It improves the anomaly localization performance without increasing the cost. To evaluate the effectiveness of the proposed method for anomaly detection and localization, extensive experiments are conducted on three widely-used anomaly detection datasets. The encouraging performance of the proposed method compared to the recent approaches based on the memory module demonstrates its superiority.

Index Terms—Anomaly detection, partition memory bank, histogram error estimation module.

I. INTRODUCTION

Visual anomaly detection aims to detect abnormal data that are different from normal visual data [1], [2], which has shown great potential in a variety of applications, such as industrial anomaly detection [3], [4], [5], [6], [7], anomaly segmentation [8], and medical diagnosis [9]. Since the abnormal patterns are diverse and the occurrence frequency of visual anomaly data is much lower than the normal ones [3], [10], [11], [12], [13], [14], it is impractical to obtain sufficient abnormal training samples with different abnormal patterns. Consequently, it is challenging to successfully detect the anomaly data by using only normal training data.

The authors are with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: xingp.ng@njjust.edu.cn; zechao.li@njjust.edu.cn).

Manuscript received 19 August 2022; revised 14 December 2022; accepted 12 January 2023. Date of publication 16 January 2023; date of current version 4 August 2023. This work was supported in part by the National Natural Science Foundation of China under Grant U20B2064 and Grant U21B2043. This article was recommended by Associate Editor D. Zeng.

Corresponding author: Zechao Li.

Some methods have been studied to address the anomaly detection task. Recently, in [17], [18], and [19], Autoencoder (AE) [20] is introduced for unsupervised anomaly detection to detect abnormal samples based on reconstruction errors [17], [21], [22]. These methods expect AE to reconstruct normal samples with only small reconstruction errors and reconstruct abnormal samples with large reconstruction errors. It is well known that AE has a strong reconstruction capability in many computer vision tasks [23], [24], [25], [26]. However, the powerful reconstruction ability of AE is not beneficial for visual anomaly detection [15], [27]. Even for unlearned anomalous samples, AE still reconstructs them well with small reconstruction errors [15]. Therefore, the reconstruction error of AE reconstruction alone is not sufficient to solve the anomaly detection challenge. Some works propose more complex self-supervised tasks, e.g., image-colorization [28]. The complex self-supervised task makes AE learn normal semantic information to recover the original image. For the unlearned abnormal samples, the abnormal semantics cannot be recovered, making the reconstruction error larger. However, with tiny differences in semantic information between abnormal and normal samples, these methods can not detect the abnormal samples quite well.

To make the reconstruction errors of abnormal samples larger than the ones of normal samples, some methods [15], [16], [27], [29] combine AE and the traditional memory modules. The output of the encoder is used as the query of a memory module, and the read of this memory module is used as the input of the decoder. Then, the memory module stores the normal features from the encoder output. When the abnormal feature queries the memory module, the output is normal feature. Therefore, the reconstruction error of abnormal sample becomes large. However, due to the query generation

Fig. 1. Illustration of the query generation methods for different memory modules. Figures (1) [15] and (2) [16] represent the logical query generated from the same position in the feature map, with different channels. Figure (3) represents the proposed query. It preserves the critical features of original feature map learned by the convolutional network.
method, the existing memory modules do not actually store normal features, but rather logical information. As shown in Figure 1 (1) and Figure 1 (2), the query generated by the traditional method is a logical reorganization of features at the same location but from different channels of the original feature map [16], [27]. A single channel feature is extracted by the same convolutional kernel by the sliding window method from the feature map of the previous layer, which contains similar critical feature information. Nevertheless, the queries generated by the logical reorganization of the feature map destroy the critical feature learned by the convolutional network, allowing the memory module to learn the reorganized logical information. When anomalous features as query address memory module, the traditional module will fit the anomalous feature reads in logical relationship to support image reconstruction. Therefore, it makes the detection and localization of anomalous samples fail with small reconstruction errors. Figure 2 (1) shows an example of detection failure for anomalous sample B, which learns a memory module by the traditional query generation method (Figure 1 (1)).

When facing tiny region anomaly samples, there is another aspect to consider - the effect of cumulative reconstruction error. Introducing the memory module makes it impossible for AE to achieve zero-error reconstruction even for normal regions. Due to the cumulative error, the abnormal scores of some normal samples can surpass the abnormal sample scores of small abnormal regions (e.g., small scratches), which leads to incorrect detection and localization results.

Moreover, Existing methods [15], [27] adopt a scheme, which concatenates the outputs of last encoder and the reads of the memory module as the input of the decoder. However, it can not well explore the storage from the memory module since encoder branch can provide sufficient features for reconstruction. Furthermore, these methods place the memory module after the last layer of the encoder focusing more on high-level features, which are not applicable for anomaly localization requiring low-level features. Consequently, all anomaly detection methods based on AE unable to reach the pixel-level anomaly localization.

Toward this end, this paper proposes a novel AE joint memory module network, as shown in Figure 3. It utilizes a Partition Memory Bank (PMB) module to improve the abilities of learning and storing normal features, as well as a Histogram Error Estimation module to adaptively eliminate the effects of cumulative reconstruction errors caused by memory module. To preserve the critical feature information learned from the original feature map, the PMB module develops a novel query generation method that utilizes each channel feature to generate a query, as shown in Figure 1 (3). It allows the memory module to learn normal semantic features rather than reorganization logical features. When anomalous features as query address the PMB module, it reads the stored normal features in a semantic relationship to get normal features. Therefore, the reconstruction error becomes larger. Figure 2(2) shows an example of a successful detection where the PMB module causes an abnormal sample B to be reconstructed as a normal sample A (the reconstruction error is large enough to be successfully detected as abnormal). Meanwhile, PMB needs to make the normal region reconstruction error as small as possible. A partition mechanism is proposed to further improve the expressiveness of the memory module, which uses multiple memory units to independently store the normal features of different partitions. Hence, PMB module enables the small reconstruction error for normal regions while the large error for the abnormal regions, which benefits the detection of the abnormal samples. The Histogram Error Estimation module utilizes the reconstruction error maps to construct histograms. Then, it is used to adaptively estimate the reconstruction error for the normal region of each image to obtain a more effective corrected error map. More importantly, it is a non-parametric method and does not consume additional resources.

To make the PMB module put more attention on low-level features for more precise localization, a new scheme is developed. We directly utilize skip connections as input of decoder on high-level features and memory module reads as input on low-level features. The reads of PMB and the outputs of the skip connections are separately utilized at different layers rather than concatenated at the same layer, which allows the memory module to learn more detailed features of normal samples. Experimental results on widely-used datasets demonstrate that the proposed method obtains competitive results.

The main contributions are summarized as follows:

1) This paper proposes a novel Partition Memory Bank (PMB) module. The unique partition mechanism and a query generation method of PMB can effectively learn and store normal features, and achieve excellent results in anomaly detection task.

2) To address the challenge of more accurate anomaly localization, a new non-parametric Histogram Error Estimation...
module is developed to eliminate the cumulative reconstruction error, which can obtain better anomaly detection results and anomaly localization maps.

(3) The AE, PMB module, and Histogram Error Estimation module are jointly explored for their optimal compatibility. Experiments are conducted on three benchmark datasets, which shows that the proposed method can effectively solve the problem of successful reconstruction of abnormal samples.

II. RELATED WORK

A. Unsupervised Visual Anomaly Detection

For unsupervised anomaly detection, the detection model is trained by only normal samples, which is used to detect abnormal samples and localize abnormal regions [18]. Traditional classifiers are introduced for anomaly detection, such as SVM [30] and OC-SVM [31]. With the popularity of deep learning, researchers proposed deep one-class methods such as DSVDD [32] and OCNN [33]. The common idea of these methods is to learn the decision boundary of normal samples. For example, DSVDD learns a spherical discriminant plane to improve discrimination efficiency. However, these methods are less effective to deal with high dimensional data.

Computer vision tasks based on pretrained models have attracted a lot of attention. Some anomaly face recognition tasks [34], [35], [36] or anomaly detection tasks [5], [37], [38] introduced pretrained models as feature extractors. In U-Std [5], the pretrained teacher network-based knowledge distill is introduced. Since the student network only learns latent representations of normal samples, the degree of sample abnormality is measured by the difference between latent feature representations from multiple student and teacher networks. However, they rely on additional training datasets and huge resource consumption. Some research proposed the employment of GAN [39] to solve the anomaly detection problem. The generator in GAN is used to learn normal sample features and discriminator to identify anomalies [40], [41], [42]. However, GAN may be unstable and tends to derive unsatisfactory results in the actual situation.

Recently, reconstruction-based methods have become popular in anomaly detection, which expects to poorly reconstruct abnormal samples and enlarges the gap of the reconstruction error of abnormal samples and normal samples [17], [18], [19], [28], [43], [44], [45], [46], [47]. For example, AE-SSIM [19] attempts to reconstruct normal samples directly using AE, while ARFAD [28] learns the reconstruction features of normal samples through self-supervised tasks such as rotation. However, these methods successfully reconstruct not only normal samples but also abnormal samples, which leads to limited performance.

Different from the above methods, this work proposes a new PMB module to reconstruct successfully normal samples and poorly abnormal samples.

B. Memory Network

The memory module was initially adapted to the field of natural language processing [23], [24], [48], [49]. The memory module includes query items and memory items as well as the results of the query. The query item and the memory item get the query weight by similarity calculation, and then read the content of the memory item by the weight to get the readings. Recently, Some existing works introduce memory bank to computer vision tasks [50], [51], such as fake video detection [52] and anomaly detection [27], [29]. Gong et al. [15] proposes MemAE, which uses a memory module to reduce the reconstruction ability of AE [20] for anomalous samples. This can be explained in terms of feature selection [53], because the memory module stores normal features and abnormal features cannot be selected from the memory module. However, due to its simple memory module and the shortcomings of the combination method, it is difficult to be effective for tiny anomaly sample detection and localization tasks. Park et al. [27] proposes a combination network of AE and memory module for video anomaly detection. It constructs an update method for the memory items of the memory module and concatenates the outputs of AE and memory module as the input of the decoder. However, it is not suitable for the localization challenges of complex scenes because it focuses
more on high-level features. Second, the output of AE as the input of the decoder results in the limitation of features stored in the memory module. Wang et al. [29] introduces an additional evaluation network on the MemAE framework as a discriminator to detect whether a sample is anomalous or not. However, it is difficult to achieve anomalous region localization with discriminators.

The most related study to the proposed method is DAAD [16]. It uses the method in Figure 1 (2) to generate memory query of size \( a \times b \times c \) \((a \leq h, b \leq w)\) and uses a discriminator to identify whether the samples are anomalous or not. DAAD uses larger query (generally \( 8 \times 8 \times c \)) to make anomalies and normals that do not share the same block pattern. However, it destroys the integrity of the features learned by the convolutional network, resulting in the memory module learning reorganized logical information. Therefore, abnormal features can also be logically fitted to anomalies such that reconstruction errors are small. The proposed PMB module employs novel query of size \( h \times w \times 1 \) (Figure 1(3)) which enables the memory module to store normal feature learned from the convolutional network. Therefore, all queries can address the PMB module to get normal features. Unlike DAAD, it is not degrade the reconstruction capability of AE, but make the memory module unable to address abnormal features. In addition, DAAD introduces discriminator with adversarial loss to solve the problem of incorrect discrimination caused by cumulative errors, which results in additional resource consumption and the failure of anomaly localization. In this paper, we not only ensure that the normal features are well reconstructed, but also the anomaly localization results can be derived from the error map without additional resources.

III. THE PROPOSED METHOD

A. Overview

This work proposes to solve the unsupervised anomaly detection problem based on the reconstruction scheme. The overview of the proposed framework is shown in Figure 3. The proposed framework contains three basic structures: 1) an AE, 2) a PMB module, and 3) a Histogram Error Estimation module.

Given the input sample \( X \), the encoder first extracts the multi-scale features \( F = \{F_1, F_2, \ldots, F_L\} \). To make full utilization of the representation information provided by the memory modules and allow the memory module to focus on low-level features, we utilize the output of the PMBs at layer 1 to layer \( p \) of the encoder and the output of the subsequent \( p+1 \) to \( L \) layers with the skip connection. Therefore, \( p \) PMBs are designed to learn and store feature maps of different scales independently. The reads of the proposed PMB module are represented by \( X_M = \{X_M^1, X_M^2, \ldots, X_M^p\} \). Then, the channel concatenation between \( X_M \) and output of previous decoder \( X_D \) as the input of the next decoder. The reconstructed image \( X_R \) is obtained after the decoder. The original difference image \( A \) is obtained by using Eq. 1,

\[
A = |X - X_R|,
\]

where \( X \) represents the input sample, and \( X_R \) represents the reconstructed sample. Besides, the Histogram Error Estimation module is developed to eliminate reconstruction errors in normal regions to obtain a more accurate corrected error map \( A' \) for anomaly localization and detection. The anomaly scores and anomaly localization maps of the samples are obtained based on \( A' \).

B. Partition Memory Bank Module

As is shown in Figure 4, The PMB module is proposed to learn and store the potential feature representations of normal samples. It achieves the challenge of anomalous sample reconstruction error becoming large by exploring the proposed joint partition mechanism and a new query generation method.

To generate the query with critical feature information, a novel query generation method is developed, as shown in Figure 1 (3). It generates queries directly from each channel
of the feature map. Thus, each query contains semantic information for a single channel, allowing the PMB module to store only normal semantic information. In addition, to enhance the learning ability and expressiveness of the memory module and to make the reconstruction error of normal regions small, a partition mechanism is proposed. It introduces multiple local units to store the normal features of different regions separately. Each memory unit is capable of storing detailed feature information for individual regions. Multiple memory units make the expressiveness of the memory module enhanced.

Specifically, for each feature map \( F_i \) with the size of \( h \times w \times c \) from \( F \), it is partitioned by using the proposed query generation approach shown in Figure 4. Along the \( 'h' \) dimension of the feature map, \( F_i \) is partitioned into \( k \) regions with the size of \( \frac{h}{k} \times w \times c \), each of which is stored and queried by a separate memory unit. The feature map of individual channel in each partition is flattened to one \( \frac{h}{k} \times w \) dimensional vector. Thus, we can get query sets \( Q = \{ q_t | t \in [1,c] \} \) in each partition. Each PMB contains \( k \) memory units \( M = \{ M_1, M_2, \ldots, M_k \} \), each memory unit contains \( m \) memory items, and the size of the memory items is the same as the query \( q_t \). Finally, as shown in Fig. 4 (4), suppose \( q_t \) is a query for \( i \)-th partition of \( F_i \), \( q_t \) and the corresponding memory items in \( M_i \) are first normalized to improve the accuracy of the attention weight. The normalization operations are as follows:

\[
\hat{q}_t = \text{norm} (q_t), \quad \hat{M}_i = \text{norm} (M_i'), \quad j \in [1,m],
\]

where \( \text{norm}(\cdot) \) represents the \( l_2 \)-norm. \( \hat{q}_t \) represents the normalized query \( q_t \), \( \hat{M}_i \) represents the memory items after normalization. Then the similarity between \( \hat{q}_t \) and the memory item \( \hat{M}_i \) is used to calculate the attention weight \( \omega_t \) as follows:

\[
\omega_t^j = \frac{\exp(\langle \hat{q}_t, \hat{M}_i^j \rangle \cdot \delta_m)}{\sum_{u=1}^{m} \exp(\langle \hat{q}_t, \hat{M}_i^u \rangle \cdot \delta_m)}
\]

Here \( \langle \cdot, \cdot \rangle \) denotes the cosine similarity. However, recent work [15] mentions that the weights \( \omega_t^j \) are too small, which leads to the results of querying the memory bank being obtained by logically fitting each memory item in the memory bank and not containing normal semantic information. Therefore, this paper proposes to filter small weights by introducing a threshold. The results of the query memory bank are only affected by the specific memory items with large weights and cannot be logically fitted by multiple memory items. It is more difficult for anomalous features to query the corresponding anomalous results from the memory bank of normal features. Therefore, when anomalous features are used as queries, only the stored normal features can be read, and then the reconstruction error of the anomalous region is large.

\[
\omega_t^j = \max \left( 0, \omega_t - \delta_m \right)
\]

where \( \delta_m \) represents the threshold. By using the memory module, we can obtain a new feature vector \( q_t' \) for each \( q_t \) by using the output of the memory unit \( M_i \).

\[
q_t' = \omega_t \cdot M_i
\]

Then a new feature map \( F' \) can be obtained by re-arranging all \( q_t' \) and fed into the decoder to generate the reconstructed image \( X_R \). It is worth noting that the stability of the reconstructed network can be improved by using \( q_t' \) since the value of \( q_t' \) does not change much.

### C. Histogram Error Estimation Module

For each image, the original error map \( A \) is obtained by using Eq. 1, in which the value of each pixel represents the reconstruction error. In the proposed approach, the reconstruction error comes from the proposed PMB module and AE. The PMB module learns and stores the normal features, which enables to guarantee that the reconstruction errors of normal regions are obviously smaller than ones of abnormal regions. However, due to the presence of the memory module, AE is difficult to achieve zero-error reconstruction in the normal region. These accumulative reconstruction errors can result in false detection from the original error map. Therefore, we need to further mitigate the impact of accumulative reconstruction errors on detection and localization performance.

In this paper, we explore a simple, yet effective Histogram Error Estimation module to estimate the error \( \mu \). As shown in Figure 3, the histogram of error map is first constructed. Due to the small reconstruction error of normal region, the pixels in normal region always lie on the left side of histogram. Therefore, we can quickly find normal pixels in the histogram. Then, a percentage \( \delta_H \) of small reconstruction errors are chosen to estimate \( \mu \) as shown in the dashed box of the histogram in Figure 3. The average value of errors of these selected pixels is utilized to estimate \( \mu \). Then, the corrected difference image \( A' \) is obtained as follows:

\[
A' = \begin{cases} 
A - \mu; & A > \mu \\
0; & A \leq \mu 
\end{cases}
\]

For anomaly detection and anomaly localization tasks, the degree of anomaly of an image is measured by the anomaly score. For the anomaly localization task, the per-pixel anomaly score \( S_T^{(i,j)} \) of each test image \( T \) needs to be obtained. For the anomaly detection task, it is necessary to obtain the overall anomaly score \( \hat{S}_T \) for \( T \). Both \( S_T^{(i,j)} \) and \( \hat{S}_T \) can be inferred from \( A' \).

\[
S_T^{(i,j)} = A_{(i,j)}, \quad \hat{S}_T = \frac{1}{h \times w} \sum_{i}^{h} \sum_{j}^{w} A'_{(i,j)}
\]

where \( h \) and \( w \) represents the height and width of the original image \( T \), respectively, while \( A'_{(i,j)} \) represents the value of position \((i, j)\) in \( A' \).

### D. Loss Function

To train the proposed model, the quality of the reconstructed image and the intuitive feeling of the human visual system are jointly explored. The reconstruction loss \( L_{MSE} \) and structural similarity index measure (SSIM) [60] loss \( L_{SSIM} \) are introduced, which are defined as:

\[
L_{MSE} = \| X - X_R \|_2
\]

\[
L_{SSIM} = \text{SSIM}(X, X_R)
\]
\[
\mathcal{L}_{\text{SSIM}} = \frac{1}{h \times w} \sum_{i=1}^{h} \sum_{j=1}^{w} (1 - \text{SSIM}^{(i,j)}(X, X_R))
\] (9)

Here \(h\) and \(w\) represent the height and width of the image. \(\text{SSIM}^{(i,j)}(X, X_R)\) represents structural similarity index between the image \(X\) and image \(X_R\) at the center \((i, j)\) position with kernel size of \(s\).

This paper proposes to jointly explore the above loss to obtain the total training loss as follows:

\[
\mathcal{L} = \mathcal{L}_{\text{MSE}} + \lambda_{\text{SSIM}} \cdot \mathcal{L}_{\text{SSIM}}.
\] (10)

where \(\lambda_{\text{SSIM}}\) is the hyperparameter of the model that measures the importance of the SSIM loss.

IV. EXPERIMENTS

This section conducts extensive experiments to verify the effectiveness of the proposed method with the comparison to several state-of-the-art methods.

A. Experimental Setting

1) Datasets: To validate the effectiveness of the proposed method, experiments are conducted on three widely-used datasets.

a) MVTec AD [3]: The MVTec AD dataset contains 5354 high-resolution color images in 15 different categories. Among them, there are 5 categories for textured images, such as “wood” or “leather”. The other 10 categories contain non-textured objects, such as “cable”. The challenge of this dataset comes from the fact that normal images and abnormal images come from the same category. The difference between the normal sample and the normal sample is subtle, such as scratches on the “leather”. Each category contains a dataset including only normal images for training and a dataset including normal and various abnormal images for testing.

b) MNIST [61]: The dataset contains 10 categories of low-resolution grayscale images. A total of 70,000 images, 60,000 pictures for training, and 10,000 pictures for testing. In anomaly detection, a category is used as a normal sample and the remaining 9 categories are used as anomaly samples (called out-of-distribution detection [62]).

c) Retinal-OCT [63]: It is a medical dataset, which contains normal (healthy) retinal CT images and 3 categories of abnormal (damaged) retinal CT images.

2) Implementation Details: In experiments, each image is resized into the size of 128 \(\times\) 128. The UNet [64]-like network is utilized as the autoencoder in this paper. The learning rate is set to 0.002. For MVTec AD, \(L\) is set to 5 and \(p\) is set to 4. The number of \(k\) in each PMB is set to \((32, 16, 16, 8)\) for five texture categories and \((16, 16, 16, 8)\) for other object categories. Accordingly, the number of memory items \(m\) is \((150, 200, 200, 200)\). For MNIST, \(L\) and \(p\) are both set to 3. The number of partitions \(k\) is set to \((1, 1, 1)\) and \(m\) is set to \((10, 10, 10)\), respectively. For Retinal-OCT, we set \(L\) and \(p\) to 4 and 3, respectively. \(k\) in each PMB is set to \((2, 2, 2)\), respectively. The number of memory items \(m\) is set to \((10, 10, 10)\), respectively. The batch size is set to 20 for the MVTec AD and Retinal-OCT datasets, as well as 64 for MNIST. For the hyper-parameters in the proposed formulation, follow [29], \(\delta_m\) is set to \(\frac{1}{m}\), \(\delta_H\) and \(\lambda_{\text{SSIM}}\) are set to 50% and 0.1, respectively.

3) Evaluation: Following [27], [41], the area under the curve (AUC) of the receiver operating characteristic (ROC) of the image level and pixel level are used to measure the performance of model.

B. Experimental Results

1) Anomaly Detection Results: The anomaly detection results on these three datasets and the comparison with the state-of-the-art methods are shown in Table I, II and III, respectively.

   a) MVTec AD: For the MVTec AD dataset, this paper compares the state-of-the-art methods with distribution-based methods and AE methods. The distribution-based approach uses pre-trained models with ImageNet dataset [69], such as CNN-Dict [54], MKDAD [37], SPADE [11], FAVA [55], Cutpaste [10], PaDiM [14]. Among them, MKDAD uses the VGG-16 [70] model pre-trained on ImageNet, SPADE and PaDiM uses the ResNet [71] pre-trained on ImageNet, and CutPaste uses the EfficientNet (B4) [72] model pre-trained on ImageNet. AE methods include CAVGA-D_\text{AE} [12], Puzzle-AE [56], ARFAD [28], memAE [15], AESc [47]. For example, Puzzle-AE introduces the complex self-supervised task of puzzle reduction, ARFAD introduces the self-supervised task of rotation and coloration, and MemAE introduces the traditional memory module.

   From Table I, the proposed method achieves mean AUROC 91.8%, which is lower than Cutpaste (96.6%) and PaDiM (95.5%) [14] that use pre-trained models. It is due to the fact that Cutpaste and PaDiM use more powerful pre-trained models to extract features, and Cutpaste also introduces faked anomalous samples to participate in training. The AE reconstruction-based anomaly detection method relies only on normal samples for training, which is an ill-posed problem. The reconstruction error of high contrast regions in normal samples is also relatively large, causing performance degradation. The semantic features extracted by the pre-trained model will better represent the anomaly patterns, which leads to the anomaly detection capability gap between the AE-based approach and the pretrained model-based approach. However, the proposed method outperforms the remaining methods using pre-trained models by 4%-10%, which validates the effectiveness of the AE combined with the PMB module. Among all AE methods, the proposed method achieves the state-of-the-art performance. The proposed approach outperforms CAVGA-D_\text{AE} that introduces real anomalous samples for training by 6%, which indicates that the proposed method performs well without relying on anomalous samples. The proposed method outperforms Puzzle-AE by 13% and ARFAD by 8%. The proposed approach that relies on image reconstruction and PMB without introducing any complex self-supervised tasks achieves the advanced performance. The proposed method outperforms the traditional memory model MemAE by 11%, which validates that the proposed PMB module learns normal features more efficiently.
TABLE I
COMPARISON OF THE PROPOSED METHOD AND ANOMALY DETECTION METHODS ON MVTec AD DATASET WITH AUROC (%). THE BEST RESULTS ARE MARKED IN BOLD. ‘✓’ MEANS THAT THE METHOD ADDITIONALLY INTRODUCES LARGE DATASETS SUCH AS IMAGENET

| Method      | Extra Datasets | Bottle | Hazelnut | Capsule | Metal | Nut | Leather | Pill | Wood | Carpet | Tile | Cable | Transistor | Toedbrush | Screw | Zipper | Mean |
|-------------|----------------|--------|----------|---------|-------|-----|---------|------|------|--------|------|-------|------------|-----------|-------|--------|------|
| Pre-trained |                |        |          |         |       |     |         |      |      |        |      |       |            |           |       |        |      |
| CNN-Dict [54] | ✓              | 78.0   | 72.0     | 84.0    | 82.0  | 87.0| 68.0    | 91.0 | 72.0 | 93.0   | 59.0 | 79.0 | 66.0       | 77.0      | 87.0  | 76.0   | 78.0 |
| MKDAD [37]  | ✓              | 99.4   | 98.4     | 80.5    | 73.6  | 95.1| 82.7    | 94.3 | 79.3 | 91.6   | 78.0 | 89.2 | 85.6       | 92.2      | 83.3  | 93.2   | 87.7 |
| SPADE [11]  | ✓              | -      | -        | -       | -     | -   | -       | -    | -    | -      | -    | -    | -         | -         | -     | -      | 85.5 |
| FAVAE [55]  | ✓              | 99.9   | 99.3     | 80.4    | 85.2  | 67.5| 82.1    | 94.8 | 67.1 | 80.5   | 97.0 | 95.0 | 93.2       | 95.8      | 83.7  | 97.2   | 87.9 |
| PaDim [14]  | ✓              | 99.9   | 92.0     | 91.3    | 98.7  | 100 | 93.3    | 99.2 | 98.8 | 98.1   | 96.7 | 92.7 | 97.4       | 96.1      | 85.8  | 90.3   | 95.5 |
| Cutpaste [10]| ✓              | 100    | 99.7     | 94.3    | 98.7  | 100 | 91.3    | 99.8 | 100  | 98.9   | 98.8 | 93.9 | 95.6       | 92.8      | 86.0  | 99.9   | 96.6 |
| AE           |                |        |          |         |       |     |         |      |      |        |      |       |            |           |       |        |      |
| CAVGA-D on [12] | ✓              | 93.0   | 90.0     | 89.0    | 81.0  | 80.0| 93.0    | 89.0 | 80.0 | 81.0   | 79.0 | 86.0 | 80.0       | 96.0      | 79.0  | 95.0   | 86.0 |
| Puzzle-AE [56] | -              | 94.2   | 91.2     | 66.9    | 66.3  | 72.9| 71.6    | 89.5 | 65.7 | 65.5   | 75.3 | 87.9 | 85.9       | 97.8      | 57.8  | 75.7   | 77.6 |
| ARFAD [28]  | -              | 94.1   | 85.5     | 68.1    | 66.7  | 86.2| 78.6    | 92.3 | 70.6 | 73.5   | 88.3 | 83.2 | 84.3       | 100.0     | 100.0 | 87.6   | 83.9 |
| MemAE [15]  | -              | 95.4   | 89.1     | 83.1    | 53.7  | 61.1| 88.3    | 95.4 | 45.4 | 63.0   | 94.6 | 69.4 | 79.3       | 97.2      | 99.2  | 87.1   | 80.2 |
| AESc [47]   | -              | 98.0   | 94.0     | 74.0    | 73.0  | 89.0| 84.0    | 95.0 | 89.0 | 99.0   | 97.0 | 89.0 | 91.0       | 74.0      | 94.0  | 89.0   |     |
| OURS        | -              | 95.2   | 99.4     | 82.3    | 84.5  | 94.5| 86.1    | 100.0| 93.1 | 97.2   | 97.1 | 85.6 | 92.4       | 95.8      | 97.0  | 77.3   | 91.8 |

TABLE II
AUROC OF THE PROPOSED METHOD AND THE COMPARED METHODS FOR ANOMALY DETECTION ON MNIST

| Method      | DSVDD [32] | CapsNetPP [57] | OCGAN [58] | LSA [59] | MemAE [15] | OCSVM [31] | ARAE [22] | OURS | MKDAD [37] | Mean |
|-------------|-------------|----------------|------------|----------|------------|------------|-----------|------|------------|------|
| Pre-trained model | -         | -              | -          | -        | -          | -          | -         | -    | -          | ✓    |
| AUROC       | 94.8        | 97.7           | 97.5       | 97.5     | 96.0       | 97.5       | 98.1     | -    | 99.35      |      |

TABLE III
AUROC OF THE PROPOSED METHOD AND THE COMPARED METHODS FOR ANOMALY DETECTION ON RETINAL-OCT

| Method      | DSVDD [32] | AnoGan [40] | VAE-GAN [65] | Cycle-GAN [66] | GANomaly [41] | P-Net [67] | MKDAD [37] | OURS | Mean |
|-------------|-------------|-------------|--------------|----------------|---------------|------------|------------|------|------|
| pre-trained model | -         | -           | -            | -              | -             | -          | -          | ✓    | -   |
| AUROC       | 74.4        | 84.8        | 90.64        | 87.39          | 91.96         | 92.88      | 97.01     | 98.00|      |

b) MNIST: For the MNIST dataset, this paper compares the state-of-the-art methods including DSVDD [32], CapsNetPP [57], OCGAN [58], LSA [59], MemAE [15], OCSVM [31], ARAE [22], and MKDAD [37]. From Table II, it can be found that the proposed method achieves mean AUROC 98.1% [62] in the out-of-distribution detection setting. The advanced performance shows the superiority of the proposed method. It is slightly lower than the pre-trained method MKDAD method but far outperforms the remain distribution methods. Besides, compared with the traditional memory module MemAE, the proposed method outperforms MemAE, which validates the effectiveness of the novel query method and partition mechanism.

c) Retinal-OCT: In the medical anomaly detection task, the proposed method achieves significant performance advantages. We compare with DSVDD [32], AnoGan [40], VAE-GAN [65], Cycle-GAN [66], GANomaly [41], P-Net [67], and MKDAD [37] on Retinal-OCT dataset. The results are shown in Table III. It can be found that the proposed method achieves the state-of-the-art performance with AUROC of 98.0%. It not only outperforms recent methods using GAN networks [65], [66], but even outperforms the MKDAD method. Our method does not rely on additional training data inductive bias specific to the pre-trained model. It is more suitable for medical image anomaly detection.

2) Anomaly Localization Results: As far as we know, none of the recent AE (including memory module) anomaly detection methods achieve pixel-level anomaly detection (anomaly localization task), which may be due to the fact that these methods weak AE reconstruction capability resulting in cannot guarantee small reconstruction error for normal region. However, the proposed method does not weaken...
TABLE IV
AUROC OF THE PROPOSED METHOD AND THE COMPARED METHODS FOR ANOMALY LOCALIZATION ON MVTec AD

| Method         | Bottle | Hazelnut | Capsule | MetalNut | Leather | Pill | Wood | Carpet | Tile | Grid | Cable | Transistor | Toothbrush | Screw | Zipper | Mean |
|----------------|--------|----------|---------|----------|---------|------|------|--------|------|------|-------|------------|------------|-------|--------|------|
| AE-SSIM [19]   | 93     | 97       | 94      | 89       | 78      | 91   | 73   | 87     | 94   | 82   | 90    | 92         | 96         | 88    | 87     |
| AE-L2 [19]     | 86     | 95       | 88      | 86       | 75      | 85   | 73   | 59     | 51   | 90   | 86    | 93         | 96         | 77    | 82     |
| AnoGAN [40]    | 86     | 87       | 84      | 76       | 64      | 87   | 62   | 54     | 50   | 58   | 78    | 90         | 90         | 80    | 74     |
| CNN-Dict [54]  | ✓      | 78       | 72      | 84       | 82      | 87   | 68   | 91     | 72   | 93   | 59    | 79         | 66         | 77    | 76     |
| MKDAD [37]     | ✓      | 96.3     | 94.6    | 95.8     | 86.4    | 98.1 | 89.6 | 84.8   | 95.6 | 82.8 | 91.8  | 82.4       | 76.5       | 96.1  | 95.9   | 93.9 |
| CAVGA-D_w [12] |        |          |         |          |         |      |      |        |      |      |       |            |            |       |        |      |
| UTAD [68]      |        |          |         |          |         |      |      |        |      |      |       |            |            |       |        |      |
| CAVGA-D_w [12] | ✓      |          |         |          |         |      |      |        |      |      |       |            |            |       |        |      |
| **OURS**       | -      | 93.7     | 92.5    | 92.1     | 85.8    | 96.7 | 91.1 | 86.5   | 92.3 | 90.7 | 94.3  | 94.2       | 80.8       | 97.5  | 97.7   | 95.4 |

TABLE V
THE RESULTS OF THE COMPARISON BETWEEN THE PROPOSED METHOD AND BASELINES. THE BEST RESULTS ARE MARKED IN BOLD. THE RESULTS OF ‘AE+skip’ ARE FROM [16]. ‘W/O H’ REPRESENTS THE PROPOSED METHOD WITHOUT HISTOGRAM ERROR ESTIMATION MODULE

| Method         | Bottle | Hazelnut | Capsule | MetalNut | Leather | Pill | Wood | Carpet | Tile | Grid | Cable | Transistor | Toothbrush | Screw | Zipper | Mean |
|----------------|--------|----------|---------|----------|---------|------|------|--------|------|------|-------|------------|------------|-------|--------|------|
| AE-SSIM [19]   | 88     | 54       | 61      | 54       | 46      | 60   | 83   | 67     | 52   | 69   | 61    | 52         | 74         | 51    | 80     | 63   |
| AE-L2 [19]     | 80     | 88       | 62      | 73       | 44      | 62   | 74   | 50     | 77   | 78   | 56    | 71         | 98         | 69    | 80     | 71   |
| AE+skip        | 71.3   | 82.8     | 74.7    | 33.6     | 57.0    | 85.3 | 97.7 | 38.5   | 98.6 | 87.9 | 57.9  | 74.9       | 74.2       | 100.0 | 69.6   | 73.6 |
| MemAE [15]     | 95.4   | 89.1     | 83.1    | 53.7     | 61.1    | 88.3 | 95.4 | 45.4   | 63.0 | 94.6 | 69.4  | 79.3       | 97.2       | 99.2  | 87.1   | 80.2 |
| DAAD [16]      | 97.5   | 89.3     | 86.6    | 55.2     | 62.8    | 89.8 | 95.7 | 67.1   | 82.5 | 97.5 | 72.0  | 81.4       | 98.9       | 100.0 | 90.6   | 84.5 |
| DAAD+ [16]     | 97.6   | 92.1     | 76.7    | 75.8     | 86.2    | 90.0 | 98.2 | 86.6   | 88.2 | 95.7 | 84.4  | 87.6       | 99.2       | 98.7  | 85.9   | 89.5 |
| **OURS (W/O H)** | 98.7   | 94.8     | 72.5    | 75.3     | 97.2    | 69.3 | 100.0| 95.0   | 88.1 | 99.1 | 81.6  | 90.8       | 87.5       | 100  | 82.4   | 88.8 |
| **OURS**       | 95.2   | 99.4     | 82.3    | 84.5     | 94.5    | 86.1 | 100.0| 93.1   | 97.2 | 97.1 | 85.6  | 92.4       | 95.8       | 97.0  | 77.3   | **91.8** |

Fig. 6. Illustrations of reconstructed images by using the proposed PMB module or the traditional memory module. ‘A’ represents normal samples and ‘B’ represents abnormal samples. ‘Ours’ shows the test results of the PMB module and ‘Trad’ shows the test results of the traditional memory module. The reconstruction error of the proposed PMB module (‘Ours’) reconstructing anomalous sample ‘B’ is very large and can be effectively detected. However, the traditional module has a small reconstruction error for abnormal samples (‘Trad’).

Compared with AE-based methods such as AE-SSIM [19] and AE-L2 [19], the proposed method gains 5%-10% improvement. Because the proposed PMB module can store the learned normal sample features, effectively making the anomalous region reconstruction error larger. Compared with the two-stage method UTAD [68], the proposed method is simpler and more effective in training, and better localization results are obtained by exploring the corrected differential image $A'$. Compared with the weakly supervised method CAVGA-D_w [12] which uses additional abnormal region annotation information, the proposed method achieves better results. It shows that the proposed method outperforms the method with weakly supervised settings in the unsupervised setting, fully validating its superiority. To further demonstrate the localization performance of the proposed method, some examples of anomaly localization are illustrated in Figure 5. It can be observed that the proposed method can well solve the problem of abnormal region localization.

C. Ablation Study

1) The Impact of the Different Memory Models: To validate the advantages of the proposed PMB module over the traditional memory module for anomaly detection, we used the OOD detection setting to demonstrate the reconstruction capability of the model but makes the anomalous features unaddresable from the PMB module thereby increasing the reconstruction error of the anomalous samples. Moreover, the unfavorable error is eliminated by using the Histogram Error Estimation module to further enlarge the difference. The results of anomaly localization on the MVTec AD dataset are shown in Table IV. As a result, our method achieves 92.1% performance in anomaly localization.
content: one category as the normal samples and the remaining categories as abnormal samples. We conducted visualization studies in two datasets (MVTec AD and MNIST) separately. The results are illustrated in Figure 6. For example, by using category ‘0’ as the normal samples for training, the samples ‘3’ and ‘4’ were reconstructed during the test. It can be observed that since the PMB module cannot successfully reconstruct the samples of class ‘3’ and ‘4’, the reconstruction results are shown as class ‘0’. The abnormal samples produce a large reconstruction error. In contrast, the traditional memory module can successfully reconstruct the class ‘3’ and ‘4’ samples with a small reconstruction error. Therefore, it is difficult to solve the anomaly detection problem by using the traditional memory module to learn the logical pixels, while the proposed PMB module can ensure the storage of normal feature information and use it for normal feature reconstruction. It can be found the similar results in the MVTec AD dataset, the PMB module has a large reconstruction error for abnormal samples.

As shown in Figure 7, we show the reconstruction results of the PMB module for more abnormal and normal samples. In the ‘Normal sample’ on the left side of Figure 7, the reconstruction error is small. In the ‘Abnormal Sample’ on the right, the reconstruction error of the anomalous region shown in the rectangular box is large, while the rest of the errors are small. It indicates that the PMB module does not weaken the reconstruction capability of AE, but makes it impossible for anomalous features to successfully address the abnormal feature output, while normal features can successfully address the normal feature output.

2) Compare With Different AE Methods and Memory Module Methods: To show the advantages of the proposed method over the reconstruction-based approach, the comparison results of different AE structures and memory modules are shown in Table V. The different baselines are set up as follows:

- **AE-SSIM**: Vanilla AE structure using SSIM loss function.
- **AE-L2**: Vanilla AE structure using L2 loss function.
- **AE+skip**: Vanilla AE structure using L2 loss function with skip connections.
- **MemAE**: Vanilla AE structure using L2 loss function with traditional memory module (Figure 1 (1)).
- **DAAD**: Vanilla AE is equipped with multi-scale block-wise memory module (Figure 1 (2)).
- **DAAD+**: DAAD+ is DAAD complemented with the adversarially learned representation.
- **OURS W/O H**: Vanilla AE is equipped with PMB module.
- **OURS**: Vanilla AE is equipped with PMB module and parameter-free Histogram Error Estimation module.

The experimental results show that the proposed method improves 27%-20% compared to the vanilla AE method, and improves compared to both MemAE and DAAD methods. These indicate the effectiveness of the novel query generation method of the PMB module, which allows the memory module to store normal features. As in the ‘WOOD’ class, AE+skip performance can reach 97.7%, while MemAE, DAAD, DAAD+ all reduce the performance, and the proposed method further improves the anomaly detection performance. Comparing the results of MemAE, DAAD, and OURS W/O H, these show that the proposed query (Figure 1 (3)) is the most effective among the three query generation methods shown in Figure 1.

3) Impact of Histogram Error Estimation Module: To verify the effect of the proposed Histogram Error Estimation module...
for anomaly detection and localization, experiments are conducted by varying $\delta_H$ within [0%, 25%, 50%, 75%, 90%]. The anomaly detection results are shown in Table VI. When $\delta_H = 0\%$, which means that the Histogram Error Estimation module is not exploited, the detection performance dramatically decreases. As $\delta_H$ is increased to 75%, the performance of anomaly detection is improved and is 2% higher than the performance without the error estimation module. When $\delta_H = 90\%$, the result of anomaly detection decreases to 89.96%. The overall results in Table VI show that the Histogram Error Estimation module can well estimate the unfavorable reconstruction error caused by AE and effectively improve the performance of anomaly detection. However, when $\delta_H$ is close to 90%, the performance of anomaly detection is decreased. The reason is anomaly region errors affecting the estimation of $\mu$, resulting in a large estimation bias in $\mu$. For anomaly localization, after eliminating the reconstruction error caused by AE, $A'$ highlights the favorable reconstruction error of the anomaly regions caused by the PMB module, which significantly improves the anomaly localization. From Figure 8, it can be found that the localization performance of all categories is greatly improved when the proposed histogram error estimation module is introduced. By comprehensively considering the performance of anomaly localization and detection, $\delta_H$ is set to 50% in experiments.

4) Impacts of $L$ and $p$: In the proposed method, the combination strategy of the proposed PMB module and the skip connection are utilized. Experiments are conducted on the MVTec AD dataset to study the sensitiveness of $L$ and $p$. The results are shown in Figure 9. From the results by varying $p$ from 3 to 5 with $L = 5$, it can be seen that the alternate utilization of the skip connection and PMB can improve the detection performance, especially for texture images. Compared with the setting ‘E’ which does not use the skip connection, the performance by using setting ‘D’ improves by 6%. Because the detailed features provided by the PMB module guarantee the reconstruction ability of samples while the skip connection provides high-level information to keep the category consistent. Comparing the results for setting ‘B’ and setting ‘E’, it shows that introducing the PMB module only at a shallow level is more effective. As we known, the deeper layers of the encoder learn semantic information, while the shallower layers learn more low-level features [73]. Therefore, the introduction of the PMB module at a shallow layer makes it possible to learn more critical detailed information that facilitates the anomaly detection task. Meanwhile, feeding high-level semantic information directly to the decoder is beneficial to ensure the consistency of semantic information and the overall image reconstruction. For anomaly localization, it is found that the larger the $p$, that is, the more PMBs are introduced, the better the localization performance. In particular, the performance changes more significantly in the texture images, indicating that multiple PMBs can learn detailed normal features more effectively for subtle anomaly localization.

5) Impact of $k$: To evaluate the effect of the number of partitions $k$ on anomaly detection, experiments with different settings were conducted on three datasets. In the MVTec AD dataset, the partitions were set to $k0$: (32, 32, 32), $k1$: (8, 8, 8, 8), $k2$: (16, 16, 16, 16), $k3$: (32, 32, 16, 16), $k4$: (32, 32, 16, 16), $k5$: (32, 16, 16), $k6$: (32, 32, 16, 16). $\delta_H$ is taken as 50%. The results of anomaly detection and localization on MVTec AD are shown in Table VII. The results of anomaly detection on the remaining datasets are shown in Table VIII. For the anomaly detection task, on the MVTec AD dataset, the larger $k$ the better the anomaly detection performance. Conversely, the smaller $k$ for MNIST and Retinal-OCT datasets, the better the detection performance. The reason is that the MVTec dataset has a large resolution and carries a large amount of information. Thus multiple storage units are adapted to store the detailed features to ensure the quality of normal samples reconstruction. While MNIST and Retinal-OCT have low resolution and simple semantic information, too large $k$ will make the anomaly features easy to generalize. Therefore, a smaller $k$ is adopted. For the anomaly localization task, the experimental results show that the smaller the $k$ value, the better the localization performance, while the larger the $k$ value, the smaller the impact on localization.

**TABLE VI**

| $\delta_H$ | 0% | 25% | 50% | 75% | 90% |
|-----------|----|-----|-----|-----|-----|
| AUROC     | 88.1 | 90.1 | 90.56 | 91.05 | 89.96 |

**Fig. 9.** The influences of the combination of skip connection and PMB on anomaly detection and anomaly localization. ‘Det’ denotes anomaly detection and ‘Loc’ denotes anomaly localization.

**TABLE VII**

| Dataset      | $k0$ | $k1$ | $k2$ | $k3$ | $k4$ | $k5$ | $k6$ |
|--------------|-----|-----|-----|-----|-----|-----|-----|
| Det          | 85.6 | 83.9 | 89.6 | 90.7 | 88.8 | 90.4 | 91.5 |
| Loc          | 93.0 | 91.1 | 91.1 | 90.8 | 89.7 | 89.1 | 89.5 |

**TABLE VIII**

| Dataset      | $k=(1,1,1)$ | $k=(2,2,2)$ | $k=(4,4,4)$ | $k=(8,8,1)$ |
|--------------|-------------|-------------|-------------|-------------|
| MNIST        | 98.1 | 96.5 | 97.6 | 95.8 |
| Retinal-OCT  | 96.8 | 98.0 | 96.8 | - |
The reason is that the expressiveness is relatively weaker when $k$ becomes smaller, and the anomalous features are less likely to be expressed as normal features. Therefore, after eliminating the errors with the histogram error estimation module, the anomaly localization performance is excellent. It is worth noting that the values of $k$ for anomaly localization and anomaly detection are not contradictory. Larger $k$ values indicate a larger number of partitions, more detailed features can be learned, and the overall reconstruction error of normal samples is smaller. Therefore, a large $k$ value learns limited normal features and better amplifies the error after reconstruction of normal and abnormal pixels. Due to the Histogram Error Estimation module, excellent localization results can still be achieved after adaptive elimination of the unfavorable error.

V. CONCLUSION

This paper proposes a new unsupervised visual anomaly detection method by jointly exploring AE and a novel memory module. To address the problem of the existing memory module, a new PMB module with a novel query generation method can learn and store the features of normal samples. With the successful reconstruction of normal features, it makes abnormal features only fit the normal features stored in PMB, which results in the reconstruction error of abnormal regions being larger. Furthermore, to eliminate the cumulative reconstruction error caused by AE, a novel Histogram Error Estimation module is proposed by exploring the reconstruction errors of normal regions. The detection and localization performance is significantly improved. Finally, we explored the optimal combination of memory module and skip connections. Extensive experiments are conducted on three widely-used datasets to verify the effectiveness of the proposed method for visual anomaly detection. In the future, we will explore a more suitable model to estimate the error caused by AE and will explore the generalizability of the proposed PMB module to other computer vision tasks.

REFERENCES

[1] I. Golan and R. El-Yaniv, “Deep anomaly detection using geometric transformations,” in Proc. NeurIPS, 2018, pp. 9781–9791.

[2] K. Leung and C. Leckie, “Unsupervised anomaly detection in network intrusion detection using clusters,” in Proc. ACCS, 2005, pp. 333–342.

[3] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, “MVTec AD—A comprehensive real-world dataset for unsupervised anomaly detection,” in Proc. CVPR, Jun. 2019, pp. 9592–9600.

[4] X. Mo, V. Monga, R. Bala, and Z. Fan, “Adaptive sparse representations for video anomaly detection,” IEEE Trans. Circuits Syst. Video Technol., vol. 24, no. 4, pp. 631–645, Apr. 2014.

[5] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, “Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 4183–4192.

[6] S. Mei, Y. Wang, and G. Wen, “Automatic fabric defect detection with a multi-scale convolutional denoising autoencoder network model,” Sensors, vol. 18, no. 4, p. 1064, Apr. 2018.

[7] S. Zhang et al., “Influence-aware attention networks for anomaly detection in surveillance videos,” IEEE Trans. Circuits Syst. Video Technol., vol. 32, no. 8, pp. 5427–5437, Aug. 2022.

[8] Y. Xia, Y. Zhang, F. Liu, W. Shen, and A. L. Yuille, “Synthesize then compare: Detecting failures and anomalies for semantic segmentation,” in Proc. ECCV, 2020, pp. 145–161.

[9] Z. Li et al., “Thoracic disease identification and localization with limited supervision,” in Proc. CVPR, Jun. 2018, pp. 8290–8299.

[10] C.-L. Li, K. Sohn, J. Yoon, and T. Pfister, “CutPaste: Self-supervised learning for anomaly detection and localization,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2021, pp. 9664–9674.

[11] J. Cohen and Y. Hoshen, “Sub-image anomaly detection with deep pyramid correspondences,” 2020, arXiv:2005.02357.

[12] S. Venkataramanan, K.-C. Peng, R. V. Singh, and A. Mahalanobis, “Attention guided anomaly localization in images,” in Proc. ECCV, 2020, pp. 485–503.

[13] Z. Li, J. Liu, C. Xu, and H. Lu, “MLRank: Multi-correlation learning to rank for image annotation,” Pattern Recognit., vol. 46, no. 10, pp. 2700–2710, Oct. 2013.

[14] T. Defard, A. Setkow, A. Loesch, and R. Audigier, “PadDM: A patch distribution modeling framework for anomaly detection and localization,” in Proc. ILCR, 2021, pp. 475–489.

[15] D. Gong et al., “Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection,” in Proc. ICCV, 2019, pp. 1705–1714.

[16] J. Hou, Y. Zhang, Q. Zhong, D. Xie, S. Pu, and H. Zhou, “Divide-and-assemble: Learning block-wise memory for unsupervised anomaly detection,” in Proc. IEEE/CVF Int. Conf. Comput. Vis., Oct. 2021, pp. 8771–8780.

[17] H. Zhai, Y. Cheng, W. Lu, and Z. Zhang, “Deep structured energy based models for anomaly detection,” in Proc. ICML, 2016, pp. 1100–1109.

[18] B. Zong et al., “Deep autoencoding Gaussian mixture model for unsupervised anomaly detection,” in Proc. ICLR, 2018.

[19] P. Bergmann, S. Löwe, M. Fauser, D. Sattlegger, and C. Steger, “Improving unsupervised defect segmentation by applying structural similarity to autoencoders,” in Proc. 14th Int. Joint Conf. Comput. Vis., Imag. Comput. Graph. Theory Appl., 2019, pp. 372–380.

[20] D. P. Kingma and M. Welling, “Auto-encoding variational Bayes,” 2014, arXiv:1312.6114.

[21] S. Wold, K. Esbensen, and P. Geladi, “Principal component analysis,” Chemometrics Intell. Lab. Syst., vol. 2, no. 4, pp. 37–52, 1986.

[22] M. Salehi et al., “ARAE: Adversarially robust training of autoencoders improves novelty detection,” Neural Netw., vol. 144, pp. 726–736, Dec. 2021.

[23] A. Kumar et al., “Ask me anything: Dynamic memory networks for natural language processing,” in Proc. ICMAL, 2016, pp. 1378–1387.

[24] C. Fan, X. Zhang, S. Zhang, W. Wang, C. Zhang, and H. Huang, “Heterogeneous memory enhanced multimodal attention model for video question answering,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 1999–2007.

[25] W. Lu, J. Zhang, X. Zhao, W. Zhang, and X. Huang, “Secure robust JPEG steganography based on AutoEncoder with adaptive BCH encoding,” IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 7, pp. 2909–2922, Jul. 2021.

[26] Z.-S. Liu, W.-C. Siu, and Y.-L. Chan, “Photo-realistic image super-resolution via variational autoencoders,” IEEE Trans. Circuits Syst. Video Technol., vol. 31, no. 4, pp. 1351–1365, Apr. 2021.

[27] H. Park, J. Noh, and B. Ham, “Learning memory-guided normality for anomaly detection,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2020, pp. 14372–14381.

[28] F. Ye, C. Huang, J. Cao, M. Li, Y. Zhang, and C. Lu, “Attribute restoration framework for anomaly detection,” IEEE Trans. Multimedia, vol. 24, no. 116–127, 2022.

[29] T. Wang, X. Xu, F. Shen, and Y. Yang, “A cognitive memory-augmented network for visual anomaly detection,” IEEE/CACI J. Autom. Sinica, vol. 8, no. 7, pp. 1296–1307, Jul. 2021.

[30] B. Schölkopf and A. J. Smola, Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond (Adaptive computation and machine learning series). Cambridge, MA, USA: MIT Press, 2002. [Online]. Available: https://www.worldcat.org/oclc/48970254

[31] Y. Chen, X. S. Zhou, and T. Huang, “One-class SVM for learning in image retrieval,” in Proc. ICIP, vol. 1, 2001, pp. 34–37.

[32] L. Ruff et al., “Deep one-class classification,” in Proc. ICLR, 2018, pp. 4393–4402.

[33] R. Chalapathy, A. Krishna Menon, and S. Chawla, “Anomaly detection using one-class neural networks,” 2018, arXiv:1802.06360.

[34] S. Ge, J. Li, Q. Ye, and Z. Luo, “Detecting masked faces in the wild with LLE-CNNs,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 2682–2690.

[35] S. Ge, C. Li, S. Zhao, and D. Zeng, “Occluded face recognition in the wild by identity-diversity inpainting,” IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 10, pp. 3387–3397, Oct. 2020.
Xing and Li: Visual Anomaly Detection Via PMB Module and Error Estimation

Zehao Li (Senior Member, IEEE) received the B.E. degree from the University of Science and Technology of China in 2008, and the Ph.D. degree from the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, in 2013. He is currently a Professor with the Nanjing University of Science and Technology. His current research interests include anomaly detection and unsupervised deep learning.

He serves as an Associate Editor for IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS.