

LogicAL: Towards logical anomaly synthesis for unsupervised anomaly localization

Ying Zhao

Ricoh Software Research Center (Beijing) Co., Ltd.

zy_deepwhite_zy@hotmail.com

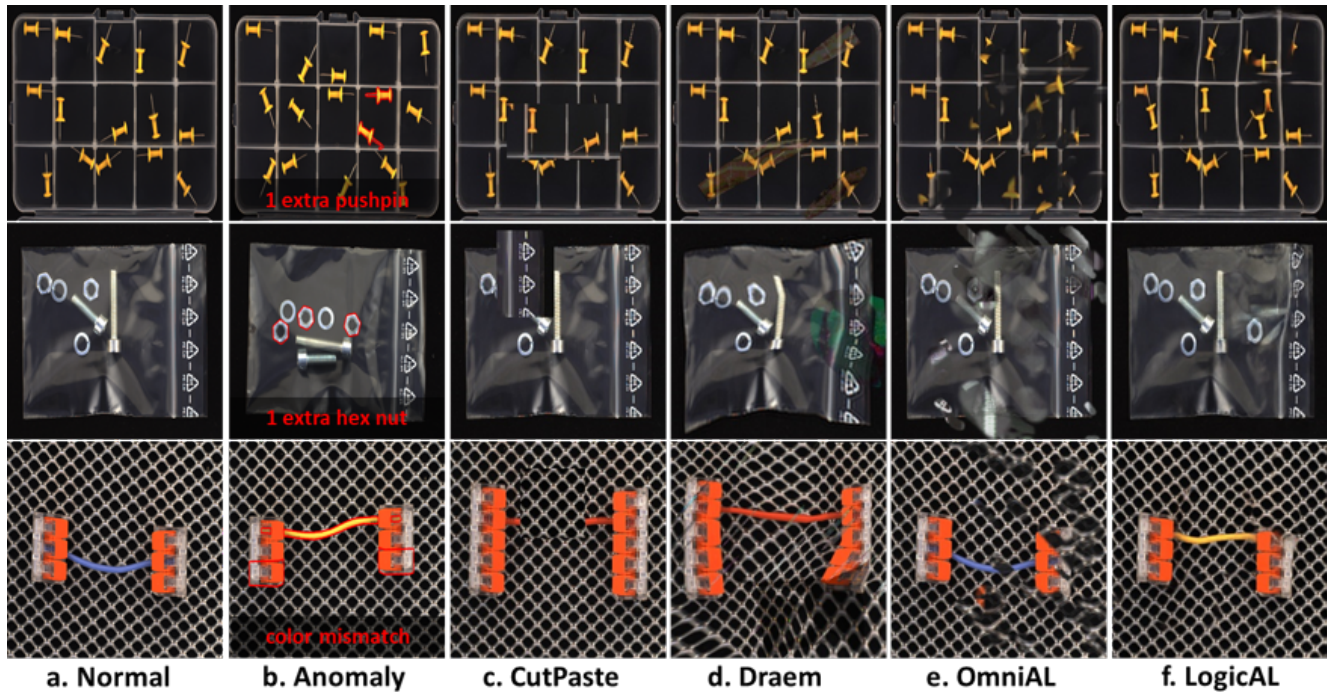


Figure 1. **LogicAL is a novel anomaly synthesis framework for unsupervised anomaly localization.** (a) and (b) are real normal and anomaly examples from MVTECLOCO [4] dataset. As shown in (f), LogicAL can generate photo-realistic anomalies that are flawless in appearance but violate the underline logic constraints, such as adding extra component and causing mismatch among components. On the contrary, the synthetic anomalies using (c) [15], (d) [31] and (e) [34] are prone to be like structural anomalies that have faulty appearance.

Abstract

Anomaly localization is a practical technology for improving industrial production line efficiency. Due to anomalies are manifold and hard to be collected, existing unsupervised researches are usually equipped with anomaly synthesis methods. However, most of them are biased towards structural defects synthesis while ignoring the underlying logical constraints. To fill the gap and boost anomaly localization performance, we propose an edge manipulation based anomaly synthesis framework, named LogicAL, that produces photo-realistic both logical and structural anomalies. We introduce a logical anomaly generation strategy that is adept at breaking logical constraints and a structural anomaly generation strategy that complements to the struc-

tural defects synthesis. We further improve the anomaly localization performance by introducing edge reconstruction into the network structure. Extensive experiments on the challenge MVTECLOCO, MVTECAD, VisA and MADsim datasets verify the advantage of proposed LogicAL on both logical and structural anomaly localization.

1. Introduction

Anomaly localization has brought a great influence on the intelligent industrial manufacturing and extremely high valuable on extensive applications. It reveals the location and degree of various anomalies that are different from the normals. Due to the anomalies are unexpected and inex-

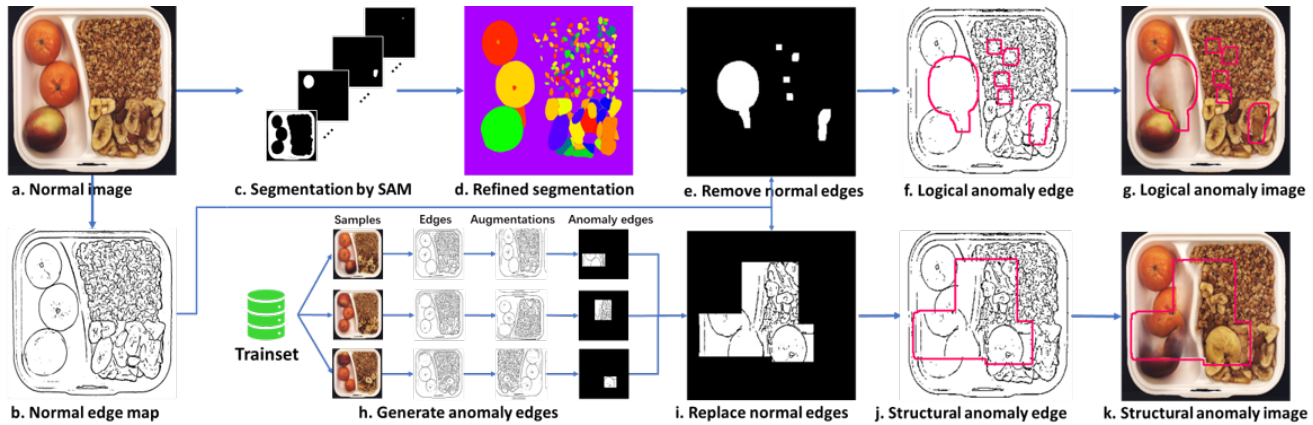


Figure 2. **Framework of LogicAL.** LogicAL converts a normal image into an anomaly image based on edge manipulation. (c) to (g) show that it generates logical anomaly image by removing normal edges from the selected semantic regions. It synthesizes structural anomaly image by replacing normal edges in arbitrary regions with the augmented edges that are sampled from trainset, as shown in (h) to (k).

haustible, it is challenging to construct a robust anomaly detector only with very limited normal data. Even though anomaly localization has spawned many superior unsupervised methods, the task is not yet completely resolved.

To handle with the shortage of real anomaly data, many appealing unsupervised approaches [15, 20, 31, 33, 34] have arose in accompany with anomaly synthesis. As examples illustrated in Fig.1c-e, existing synthesis methods [15, 31, 34] can only randomly destroy the appearance of normals shown in Fig.1a[4] and produce the anomalies having different visual structures with the normals. Fig.1b shows some more challenging cases that are beyond the capabilities of these synthesis methods. Each component in Fig.1b is flawless but their combination violates the certain underlying logical or geometrical constraints of the normal data. The anomaly detectors trained only with the synthetic structural anomalies would be vulnerable to logical anomalies.

By comparing the difference between real logic anomaly (Fig.1b) and the synthetic anomaly (Fig.1d-e), we can see that existing synthesis methods [31, 34] only modify discrete local regions and ignore the long-range dependencies. It leads to a more structure-oriented defects generation. Moreover, the fake boundaries introduced by a naive large region replacement make the anomaly detector learns a simple decision. Therefore, a photo-realistic logical anomaly synthesis method for unsupervised anomaly localization is highly demanded.

To conquer aforementioned problems, we propose a novel anomaly synthesis and localization method LogicAL for effectively localizing anomaly pixels of both logical and structural anomalies. Fig.2 illustrates overall concept of LogicAL. To generate photo-realistic anomalies, it manipulates edges and converts the modified edge map into image by using edge-to-image generator. To generate both logical and structural anomalies as shown in Fig.2g and Fig.2k, LogicAL modifies the normal edge map by using differ-

ent region selection and edge manipulation strategies. By using pre-trained segmentation network SAM [14], LogicAL can manipulate the edges from a semantic region. As shown in Fig.2c to Fig.2g, it synthesizes a logical anomaly status that the breakfast-box is missing one orange. Fig.2h to Fig.2k demonstrate the structural anomaly generation. To forge structural anomaly edges, LogicAL replaces the normal edges from arbitrary regions with the augmented (flip or resize) edges that are sampled from the trainset. Beside removing and replacing normal edges, LogicAL also uses strategy of merging edges to synthesizes more complex situations of mixing logical and structural anomalies, as shown in Fig.3. Fig.4 shows the anomaly localization framework that equipped with LogicAL anomaly synthesis. It produces high quality reconstruction and precise anomaly localization for both logical and structural anomalies.

In summary, we make following main contributions:

- We propose a logical anomaly synthesis method that generates photo-realistic anomaly images violating underlying logical constraints. It fills the gap left by existing structure-oriented anomaly synthesis methods.
- We introduce different edge manipulation and region selection strategies to balance logical and structural anomaly synthesis. By using the proposed anomaly synthesis, we improve existing reconstruction-based methods without effecting their inference runtime.
- We further improve the anomaly localization performance by introducing edge reconstruction into the network structure. We achieve superior anomaly localization performance, 69.7 pixel AU-sPRO and 98.3 pixel AU-ROC, on challenge MVTecLOCO [4] and MVTecAD [2] datasets respectively.

2. Related works

Anomaly synthesis. Anomaly synthesis is closely paired with unsupervised anomaly localization methods [10, 12,

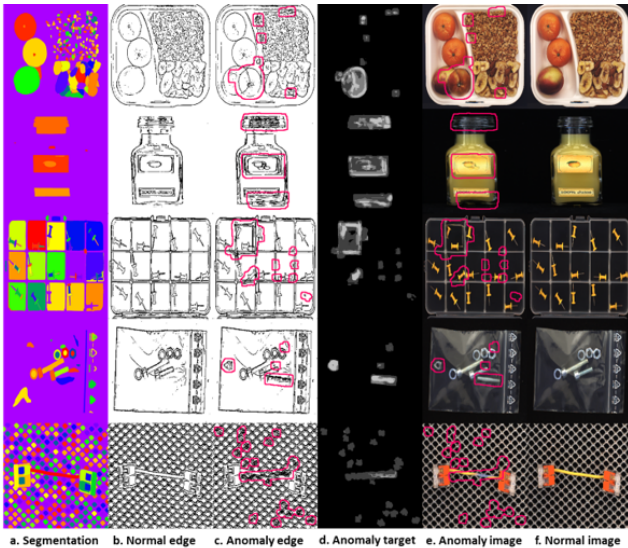


Figure 3. **Visualization of our synthetic anomalies.** (a) and (b) are extracted from (f). (c) is generated by merging the normal edges (b) from the selected regions of (a) with sampled edges. Anomaly image (e) is converted from (c). (d) is the difference map between (e) and (f).

[15, 31–33] due to the lack of real anomaly samples. CutPaste [15] propose a simple but effective method that synthesises anomaly by cutting a local rectangular region from the normal image and paste it back at a random position. It learns representations by classifying normal data from the synthetic anomalies. SPD [36] uses a smoothed version of CutPaste [15] augmentation. Extensive experiments from Draem [31] show that the quality of synthetic anomaly highly effects the performance. Instead of using simple regular shaped anomalous, Draem [31] carefully designs the anomaly regions mask by using binarized Perlin noise. To increase the anomaly diversity, it extracts the anomaly texture from an extra DTD [5] dataset. To simulate photo-realistic anomaly samples, JNLD [33] proposes a multi-scale noticeable anomalous generation method based on just noticeable distortion [27]. OmniAL [34] further improves the anomaly synthesis of JNLD [33] by using a panel-guided strategy to train a unified detector for N classes. Different with our method, these existing methods pay more attention on structural anomaly while ignore the logical anomaly.

To synthesises logical anomaly, SLSG [29] extends the anomaly synthesis of Draem [31] by controlling the parameters of Perlin noise and the binarization threshold to get a more concentrated anomaly region mask. More recent road anomaly detection methods [17, 28] re-synthesize the input image from the predicted semantic map using a generative adversarial network. Synboost [9] presents a pixel-wise anomaly detection framework that uses uncertainty maps to improve over existing re-synthesis methods

[17, 28] in finding dissimilarities between the input and generated images. However, these synthesis anomalies in road scene often introduce large difference in the normal regions. Inspired by existing methods, we propose a novel edge-controlled anomaly synthesis method that generates photo-realistic both logical and structural anomalies.

Anomaly localization. Anomaly localization aims to segment out the pixel-level anomaly regions. On the other hand, anomaly detection refers to distinguishing anomalous images at the image-level from the majority of normal images. Existing methods solve these two tasks either by a distance-based [3, 6, 8, 15, 21] or reconstruction-based [3, 29, 31, 33, 34] way. The distance-based methods usually rely on feature extractors pre-trained on the ImageNet. CutPaste [15] uses GradCAM [22] to get pixel-level anomaly location. SPADE [6] relies on K nearest neighbors of pixel-level feature pyramids extracted by pre-trained deep features. It detects anomaly based on alignment between an anomalous image and a constant number of the similar normal images. Instead of using time-costly clustering, PaDim [8] uses a well-known Mahalanobis distance metric [19] as an anomaly score. PatchCore [21] combines patch-level embeddings from ImageNet models with an outlier detection model. The reconstruction-based methods [29, 31, 33, 34] use segmentation sub-network to predict the defective regions by comparing the difference between reconstructed images with input.

Student-teacher (S-T) [3] network uses three different pre-trained CNN as teachers with different receptive fields. It trains three student networks to mimic their corresponding teachers on the normal images. The anomaly is revealed by the failure mimic on unseen anomalous images. To better detect both logical and structural anomalies, GCAD [4] introduces the global and local branches using S-T networks [3]. The final prediction is obtained by merging results of the two branches. SLSG [29] uses a generative pre-trained network to learn the feature embedding of normal images. To model position information in images, it uses a self-supervised task to learn the reasoning of position relationships and use the graph convolutional network to capture across-neighborhood position relationships. This paper differs from these previous works by learning edge information from the edge-controlled synthetic anomaly.

3. Methods

As shown in Fig.4, the proposed method LogicAL consists of anomaly synthesis and anomaly localization. The anomaly synthesis module yields photo-realistic anomalies by globally and locally edge control. It extracts edge with a pre-trained edge detector and converts the modified edge into anomaly image by a pre-trained generator. During training, the anomaly synthesis module alternatively provides logical and structural anomalies to the anomaly local-

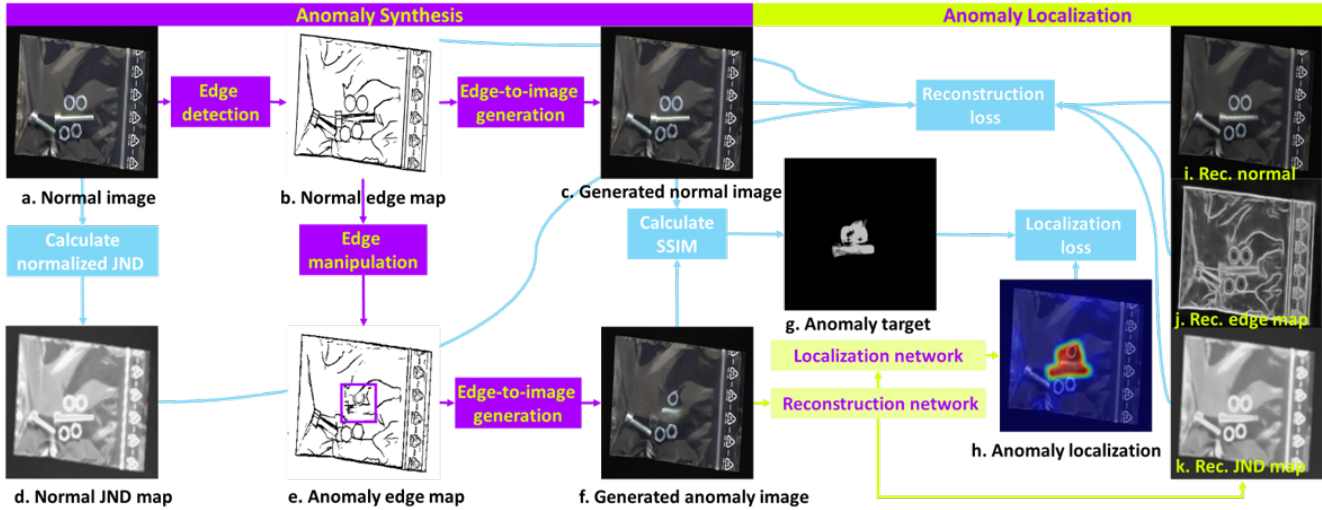


Figure 4. **Framework of proposed unsupervised anomaly localization.** It consists of anomaly synthesis and localization modules. Anomaly synthesis is based on anomaly edge map construction and edge-to-image generation. Synthetic anomaly is reconstructed into normal image, corresponding just noticeable distortion (JND) and edge maps. Anomaly localization is achieved by exploring difference between reconstructed and original data.

ization module in real-time. The anomaly localization module learns to reconstruct various normal information and spot the difference between the reconstructed and original data. In testing phase, it only takes the anomaly localization module to derive the pixel-wise prediction.

3.1. Anomaly synthesis

Edge detection. To generate photo-realistic anomalies, it is better to tune an intermediate representation rather than directly modify the source image. Edge reveals sharp brightness, color or texture changes in any part of the image. Due to its unique object-agnostic property, edge is a suitable option to be the intermediate representation. With the development of deep learning, edge detection changes from the traditional gradient-based method to the end-to-end learning-based method. PiDiNet [23] is one of the appealing edge detectors that achieve a better trade-off between accuracy and efficiency. It integrates the advantages of traditional edge detectors and deep CNNs by using the pixel difference convolution. By using PiDiNet [23], we achieve zero-shot edge detection for following edge control based anomaly synthesis. As shown in Fig.4, the extracted useful edge map Fig.4b is also used in training the anomaly localization module to avoid blurry reconstruction and inaccurate localization.

Edge-to-image generation. As shown in Fig.5, our edge-to-image generation module converts the extracted edge maps to color images with pixel-to-pixel correspondence. It is trained as the pix2pixHD [25], a conditional generative adversarial network (cGAN), with pairs of edge maps and color images from the normal dataset. The pix2pixHD [25] consists of a coarse-to-fine generator translating edge maps to color images and a multi-scale discriminator distinguishing real images from the generated ones.

It aims to model the conditional distribution of real images given the input edge maps via the minimax game. Considering the need of generation from anomaly edge, we follow DeepSIM [24] to augment the training pairs by using thin-plate-spline (TPS) [11] warps. The TPS augmentation simultaneously manipulates both edge map and color image by randomly shifting 3x3 control points in the horizontal and vertical directions. The warp is smoothed by further optimization [11]. Fig.5c-d show examples of TPS warped edge map and corresponding color image. By training with these smooth warped pairs, the generator is less likely to produce a global collapsed result when it handles the synthetic or real anomaly edges.

Edge manipulation. Given the pre-trained edge-to-image generator, it is supposed to generate anomaly images from the modified edge maps. Even though the generator is trained with carefully augmented data, it still produces global collapsed results if the input edge maps are out-of-distribution. Supplementary demonstrates that a suitable portion of anomaly edges is the key to generate high quality anomaly images. Therefore, our edge manipulation involves two region selection and three edge modification strategies.

As illustrated in Fig.2, the region selection strategies consist of semantic and semantic-agnostic regions selection. We use the pre-trained segmentation model SAM [14] to get rough semantic regions. Due to SAM [14] is over-segmented, we further refine the segmentations by removing background, grouping small regions and merging overlapped regions and build a map of candidate regions, as shown in Fig.2d. For the semantic-agnostic regions selection, we randomly combine three regions with different aspect ratios or shapes.

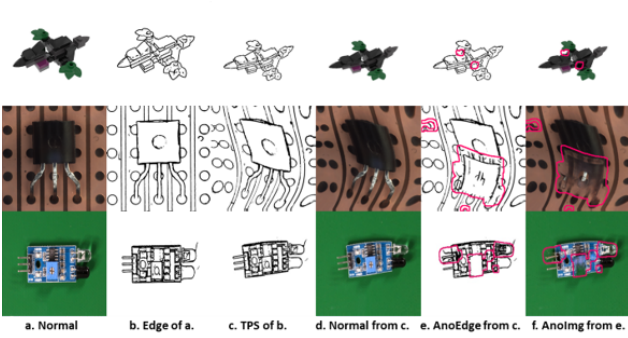


Figure 5. **Edge-to-image generation.** (a) Normal images from MADsim [35], MVTecAD [2] and VisA [36], (b) Edge maps extracted by PiDiNet [23], (c) TPS [11] warping on (b), (d) Images generated from (c) by DeepSIM [24], (e) and (f) are synthesis anomaly edge maps and images applied same TPS warping.

Given the selected regions, we modify the normal edges by removing, replacing or merging strategies. The candidate anomaly edges are derived from the normal edges belong to the same class. By doing these edge modifications, it is easier to yield the logical anomalies that violate the constraints of amount, position, matching, etc. Fig.3 shows examples of synthetic anomalies for MVTecLOCO [4] dataset. As the ambiguity exists between structural and logical anomalies, different strategies are not always strictly correspond to either logical or structural anomalies. These modifications also can fake the structural anomalies, such as scratches, dents, or contaminations.

Given the j -th normal image I_j , the corresponding anomaly image I_{A_j} is generated as follow:

$$E_{A_j} = \begin{cases} M_j + (1 - M_j) * E_j, \text{remove} \\ M_j * A(E_i) + (1 - M_j) * E_j, \text{replace} \\ M_j * (A(E_i) + E_j - 1) + (1 - M_j) * E_j, \text{merge} \end{cases} \quad (1)$$

where E_j is the source normal edge map extracted from I_j , E_i is the candidate anomaly edge map extracted from the i -th normal image I_i , A indicates random augmentation applied to E_i , the source anomaly edge map E_j is the combination of the augmented candidate anomaly edge maps $A(E_i)$, M_i is the selected regions mask. As shown in Fig.5e and Fig.5f, the synthesis anomaly edge map and image can be further augmented by applying TPS [11] warping.

3.2. Anomaly localization

Overall, our anomaly localization module follows the network structure of OmniAL [34] that consists of reconstruction and localization sub-networks equipped with dilated channel and spatial attention (DCSA) blocks and DiffNeck module. The normalized just noticeable distortion (JND) map [27, 33, 34] reveals a perceptual threshold of intensity change in an image that can be noticed by the human vision system. For both JND map and normal image recon-

struction, we not only use MSE loss to supervise the pixel-to-pixel recovering but also the structural similarity (SSIM) [26] loss to yield plausible local consistency. Different from OmniAL [34], we introduce edge information to the learning flow. In addition to learn to reconstruct the normal image and JND map, the reconstruction sub-network also produces the edge map. Following PiDiNet [23], we adopt the annotator-robust loss function proposed in [18] for the reconstructed edge maps. To localize anomaly regions more precisely, the localization sub-network calculate the reconstruction error with the assistance of edge maps.

To train the localization sub-network, we can construct the ground truth maps by comparing the synthetic anomaly images with either the source normal images or the generated images from the normal edges. Due to the error of edge-to-image generator, the normal regions in the generated and the original images are not exactly same. It is impossible to localize anomaly region simply from the generated images. The SSIM [26] provides a measure of the similarity by comparing two images based on luminance similarity, contrast similarity and structural similarity information. To avoid constructing an over-sensitive anomaly detector, we use SSIM to calculate the ground truth maps with both generated normal and anomaly images. To handle the unbalance of normal and anomaly, we use focal loss [16] to supervise the predicted anomaly localization.

4. Experiments

To demonstrate the effectiveness of proposed LogicAL, we conduct extensive experiments on the challenging MVTecLOCO [4], MVTecAD [2], VisA [36] and MADsim [35] datasets. We evaluate the overall performance of anomaly detection and localization comparing with existing advanced methods. Our experiments also include ablation studies on the effectiveness of each component.

4.1. Datasets

MVTecLOCO [4] dataset contains five categories of approximately 3,644 images covering both logical and structural anomalies from industrial inspection scenarios. Considering the ambiguity of the logical anomaly determination at the pixel level, MVTecLOCO [4] introduces a performance metric, saturated per-region overlap (sPRO), that takes the different modalities of the defects present in the dataset into account. It is a generalized version of the per-region overlap (PRO) metric that saturates once the overlap with the ground truth exceeds a certain saturation threshold. To evaluate pixel-level anomaly localization performance, we calculate the AU-sPRO score based on the area under the FPR-sPRO curve up to the false positive rate is 5%. The performance of image-level classification is evaluated by the AU-ROC.

Table 1. Performance comparison of LogicAL with existing methods on MVTecLOCO[4].

Category	Methods W/O anomaly synthesis				Methods With anomaly synthesis			
	Patchcore[21]	SINBAD[7]	GCAD[4]	EfficientAD[1]	Draem[31]	SLSG[29]	OmniAL[34]	LogicAL
	Image-level detection AU-ROC / Pixel-level localization AU-sPRO (FPR 5%)							
BreakfastBox	81.3 / 46.0	92.0 / -	83.9 / 50.2	- / -	80.3 / -	88.9 / 65.9	75.9 / 46.5	85.4 / 46.8
JuiceBottle	95.6 / 71.0	94.9 / -	99.4 / 91.0	- / -	94.3 / -	99.1 / 82.0	99.5 / 87.7	98.5 / 91.3
Pushpins	72.3 / 44.7	78.8 / -	86.2 / 73.9	- / -	68.6 / -	95.5 / 74.4	79.6 / 59.6	87.4 / 81.3
ScrewBag	64.9 / 52.2	85.4 / -	63.2 / 55.8	- / -	70.6 / -	79.4 / 47.2	83.1 / 53.2	82.0 / 52.3
SConnector	82.4 / 58.6	92.0 / -	89.3 / 79.8	- / -	85.4 / -	88.5 / 66.9	88.1 / 69.1	89.0 / 76.3
Mean	79.3 / 54.5	86.8 / -	83.3 / 70.1	90.7 / 79.8	79.8 / -	90.3 / 67.3	85.0 / 63.0	88.5 / 69.7

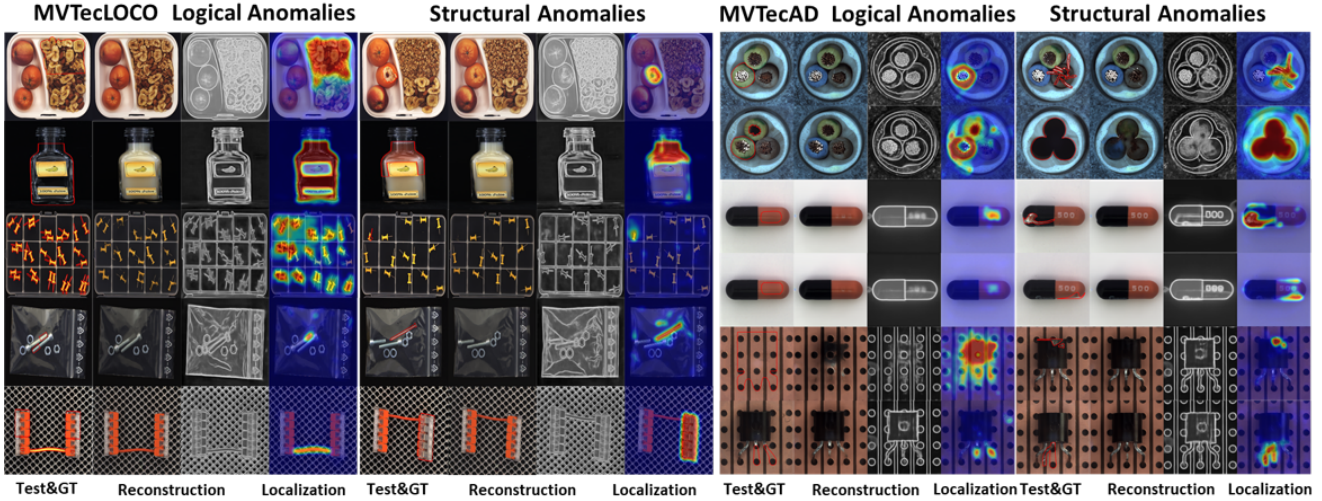


Figure 6. **Qualitative illustration of our anomaly detection results.** From left to right, the test images overlap with red ground truth contours, the reconstructions and the anomaly localization map overlays are shown in columns. The logical and structural anomalies of MVTecLOCO [4] and MVTecAD [2] are shown in groups.

Table 2. Performance comparison for both structural and logical anomalies on MVTecLOCO[4].

Method	Logical	Structural	Mean
	Image AU-ROC / Pixel AU-sPRO (FPR 5%)		
S-T[3]	66.4/49.7	88.3/75.6	77.3/62.6
SPADE[6]	70.9/53.6	66.8/36.8	68.9/45.1
GCAD[4]	86.0/ 71.1	80.6/69.2	83.3/ 70.1
SLSG[29]	89.6/-	91.4/-	90.3/67.3
LogicAL	84.6/68.8	93.6/81.5	88.5/69.7

MVTecAD [2] dataset contains 1,258 test images but pays more attention on structural anomalies than MVTecLOCO [4]. It only includes 37 test images from three out of fifteen categories matching the definition of logical anomalies, including cable, capsule and transistor. As shown in Fig.6, the logical anomalies consists of cable swap, faulty imprint and transistor misplaced. Follow [8, 29, 31, 33, 34], we evaluate our method on anomaly detection and localization with image and pixel AU-ROC.

VisA [36] dataset consists of 9,621 normal and 1,200 anomalous color images covering 12 objects in 3 domains, including complex structure, multiple instances and single

Table 3. Ablation study on MVTecLOCO [4]. Semantic: regions are extracted from refined SAM [14] segmentation, Arbitrary: random regions, Remove/Replace/Merge edges in the selected regions, TPS: apply TPS warping to synthetic anomaly edge maps and images, I: Image-level classification AUROC, P: Pixel-level localization AU-sPRO.

Region selection		Edge modification			Aug	Performance	
Semantic	Arbitrary	Remove	Replace	Merge	TPS	I	P
-	-	-	-	-	-	85.2	63.2
-	✓	-	✓	-	-	87.6	67.4
✓	-	-	✓	-	-	89.0	68.6
✓	-	✓	-	-	-	89.1	68.1
✓	-	-	-	✓	-	88.5	67.8
✓	-	✓	✓	-	-	88.7	67.9
✓	-	✓	✓	✓	-	88.5	67.0
✓	✓	✓	✓	✓	-	88.5	69.7
✓	✓	✓	✓	✓	✓	89.2	67.2

instances. The anomalous images contain various flaws, including surface defects such as scratches, dents, color spots or crack, and logical defects like misplacement or missing parts. There are 5-20 images per defect type and an image may contain multiple defects.

MADsim [35] dataset contains 5,231 normal and 4,902 anomaly color images of 20 types of 3D LEGO animal models from different viewpoints covering a wide range of poses. It uses Blender in combination with Ldrew (LEGO

Table 4. Performance comparison of LogicAL with existing methods on MVTecAD [2]. OmniAL+ indicates that the OmniAL network is trained with our synthetic anomalies.

Category	Methods W/O anomaly synthesis		Methods With anomaly synthesis				
	Patchcore[21]	EfficientAD[1]	Draem[31]	SLSG[29]	OmniAL[34]	OmniAL+	LogicAL
	Image AU-ROC / Pixel AU-ROC						
bottle	100/98.6	-/-	99.1/99.1	99.4/99.1	99.4/99.0	100/98.2	100/95.9
cable	99.5/98.4	-/-	94.7/94.7	98.3/97.4	97.6/97.1	97.6/97.2	99.0/97.6
capsule	98.1/98.8	-/-	98.5/94.3	95.5/95.9	92.4/92.2	95.5/97.2	95.5/98.2
carpet	98.7/99.0	-/-	95.5/95.5	99.0/96.0	99.6/99.6	97.4/99.3	98.5/99.5
grid	98.2/98.7	-/-	99.9/99.7	100/98.5	100/99.6	99.7/99.5	99.4/99.4
hazelnut	100/98.7	-/-	100/99.7	99.5/97.8	98.0/98.6	99.9/98.9	98.5/98.6
leather	100/99.3	-/-	100/98.6	100/99.5	97.6/99.7	100/99.8	98.1/99.5
metal nut	100/98.4	-/-	98.7/99.5	100/98.9	99.9/99.1	98.7/96.6	99.4/97.6
pill	96.6/97.4	-/-	98.9/97.6	99.2/98.0	97.7/98.6	97.4/98.5	97.5/99.1
screw	98.1/99.4	-/-	93.9/97.6	89.1/97.3	81.0/97.2	100/98.9	99.3/99.3
tile	98.7/95.6	-/-	99.6/99.2	100/98.6	100/99.4	100/98.4	100/99.0
toothbrush	100/98.7	-/-	100/98.1	100/99.4	100/99.2	99.7/99.6	100/99.4
transistor	100/96.3	-/-	93.1/90.9	97.3/92.5	93.8/91.7	95.0/91.0	98.2/97.3
wood	99.2/95.0	-/-	99.1/96.4	99.6/96.8	98.7/96.9	94.1/94.3	99.0/93.9
zipper	99.4/98.8	-/-	100/98.8	100/97.1	100/99.7	100/99.7	100/99.5
Mean	99.1/98.1	99.1/-	98.0/97.3	98.5/97.5	97.0/97.8	98.3/97.8	98.8/98.3

Table 5. Performance comparison of LogicAL with existing methods on VisA[36]. OmniAL+ indicates that the OmniAL network is trained with our synthetic anomalies. I-AUC: Image AUROC, P-AUC: Pixel AUROC, P-AP: Pixel AP.

Category	Padim+SPD[36]		Draem[31]			JNLD[33]			OmniAL[34]			OmniAL+			LogicAL			
	I-AUC	P-AUC	I-AUC	P-AUC	P-AP	I-AUC	P-AUC	P-AP	I-AUC	P-AUC	P-AP	I-AUC	P-AUC	P-AP	I-AUC	P-AUC	P-AP	
Complex structure	PCB1	92.7	97.7	71.3	98.6	60.4	82.0	96.4	72.8	96.6	98.7	63.5	97.3	99.2	62.4	95.0	99.0	72.1
	PCB2	87.9	97.2	89.7	92.5	3.5	96.3	91.9	31.2	99.4	83.2	2.8	99.5	94.3	33.3	97.8	93.9	31.4
	PCB3	85.4	96.7	73.1	93.8	18.7	96.9	95.3	43.4	96.9	98.4	56.9	98.4	99.6	62.7	94.1	98.9	52.0
	PCB4	99.1	89.2	91.3	95.8	32.7	94.8	96.1	37.4	97.4	98.5	38.4	98.0	98.6	49.0	97.4	98.6	50.5
Multiple instances	Macaroni1	85.7	98.8	70.3	95.8	8.2	94.3	98.8	25.9	96.9	98.9	7.6	95.9	97.4	17.4	94.0	99.0	8.5
	Macaroni2	70.8	96.0	71.3	94.1	25.4	86.5	92.9	17.2	89.9	99.1	11.4	85.0	96.9	21.3	86.6	99.1	20.9
	Capsules	68.1	86.3	77.3	93.7	20.2	89.1	98.9	38.6	87.9	98.6	62.9	91.6	99.3	61.4	93.9	99.7	65.1
	Candles	89.1	97.3	82.3	87.0	27.9	89.1	94.8	25.9	85.1	90.5	29.2	89.6	93.3	20.0	92.4	90.2	23.6
Single instance	Cashew	90.5	86.1	94.2	94.7	41.2	96.0	96.3	43.7	97.1	98.9	77.3	97.3	97.6	81.1	97.9	97.6	66.3
	Chewing gum	99.3	96.9	93.4	97.5	40.9	98.5	99.4	74.7	94.9	98.7	82.9	96.1	99.4	80.6	98.5	99.5	73.8
	Fryum	89.8	88.0	100	97.5	40.9	93.2	95.8	42.9	97.0	89.3	28.3	96.8	95.1	31.0	99.0	93.7	29.2
	Pipe fryum	95.6	95.4	94.1	81.8	23.7	96.0	97.0	44.1	91.4	99.1	69.1	97.1	98.6	70.4	97.0	99.5	72.9
Mean	87.8	93.8	84.1	88.8	25.4	93.0	96.1	41.5	94.2	96.0	44.2	95.2	97.5	49.2	95.3	97.4	47.2	

parts library) to build three types of anomalous, such as burrs, stains and missing parts. To evaluate pose-agnostic anomaly detection ability, we also evaluate our method on MADsim dataset. Fig.5 illustrates examples from MADsim [35], MVTecAD [2] and VisA [36] datasets.

4.2. Implementation

We extract four scales edge maps with pre-trained PiDiNet [23]. Since the first scale edge map contains more detail edges, we further use it to train edge-to-image generators and anomaly localization network. With the pairs of extracted edge map and color image, we firstly train a unified DeepSIM [24] model 300 epochs with a batch size of 56 images having size of 256x256 for each dataset. Different edge manipulation strategies are used to synthesis anomaly edge maps alternatively. The pre-trained DeepSIM models are used to convert synthesis anomaly edge maps to anomaly images on-the-fly during training the anomaly localization network. For anomaly localization, we train the network 300 epochs with a batch size of 15(12) images having size of 256x256(256x320) for MVTecAD [2], MVTeCLOCO [4] and MADsim [35](VisA [36]). The Adam optimizer has an initial learning rate of 1e-4 and decreases the

learning rate with multi-step schedule.

4.3. Performance

Quantitative evaluation. Table 1 and 2 show our quantitative comparison with the state-of-the-art methods recently reported on MVTeCLOCO [4] for both logical and structural anomaly detection and localization. Performance of the other methods in Table 1 are reported from SLSG [29]. Separate performance comparison of logical and structural anomalies are shown in Table 2. Overall, we achieve better comprehensive performance of anomaly detection (88.5 AU-ROC) and localization (69.7 AU-sPRO) than existing methods. Even though SLSG [29] achieves good performance in anomaly detection with 90.3 AU-ROC, it only has 67.3 AU-sPRO score in anomaly localization. The similar situation happens to GCAD [4] that is proposed along with the MVTeCLOCO dataset. Table 4 shows our superior performance comparing with anomaly synthesis based methods on the MVTecAD dataset. By training with our synthetic anomalies, we improve OmniAL [34] by 1% performance on image-level anomaly detection. With edge reconstruction, our method LogicAL achieves further 0.5% improvement on both anomaly detection and localization.

Table 6. Performance comparison of LogicAL with existing methods on MADsim [35].

Category	Methods W/O anomaly synthesis				Methods With anomaly synthesis		
	Patchcore[21]	CFlow[13]	UniAD[30]	PAD[35]	Cutpaste[15]	Draem[31]	LogicAL
	Image AU-ROC / Pixel AU-ROC						
Gorilla	66.8/88.4	69.2/94.7	56.6/93.4	93.6/99.5	-/36.1	58.9/77.7	58.2/ 91.7
Unicorn	92.4/58.9	82.3/89.9	73.0/86.8	94.0/98.2	-/69.6	70.4/26.0	77.0/82.4
Mallard	59.3/66.1	74.9/87.3	70.0/85.4	84.7/97.4	-/40.9	34.5/ 47.8	80.0/33.2
Turtle	87.0/77.5	51.0/90.2	50.2/88.9	95.6/99.1	-/77.2	18.4/45.3	84.0/90.1
Whale	86.0/60.9	57.0/89.2	75.5/90.7	82.5/98.3	-/66.8	65.8/55.9	74.2/86.7
Bird	82.9/88.6	75.6/91.8	74.7/91.1	92.4/95.7	-/71.7	69.1/60.3	71.7/92.9
Owl	72.9/86.3	76.5/94.6	65.3/92.8	88.2/99.4	-/51.9	67.2/78.9	53.5/ 88.0
Sabertooth	76.6/69.4	71.3/93.3	61.2/90.3	98.5/95.7	-/71.2	68.6/26.2	70.2/86.7
Swan	75.2/73.5	67.4/93.1	57.5/90.6	98.8/86.5	-/57.2	59.7/75.9	64.3/91.7
Sheep	89.4/79.9	80.9/94.3	70.4/92.9	90.1/97.7	-/67.2	59.5/70.5	79.5/93.5
Pig	85.7/83.5	72.1/97.1	54.6/94.8	88.3/97.7	-/52.3	64.4/65.6	64.9/92.4
Zalika	68.2/64.9	66.9/89.4	50.5/86.7	88.2/99.1	-/43.5	51.7/66.6	52.9/83.4
Phoenix	71.4/62.4	64.4/87.3	93.8/91.7	82.3/99.4	-/53.1	53.1/38.7	59.7/83.9
Elephant	78.6/56.2	70.1/72.4	55.4/84.7	92.5/99.0	-/56.9	62.5/55.9	57.0/ 82.3
Parrot	78.0/70.7	67.9/86.8	53.4/85.6	97.0/99.5	-/55.4	62.3/34.4	57.1/ 83.9
Cat	78.7/85.6	65.8/94.7	53.1/93.8	84.9/97.7	-/58.3	61.3/79.4	64.6/95.4
Scorpion	82.1/79.9	79.5/91.9	69.5/92.2	91.5/95.9	-/71.2	83.7/79.7	78.9/ 92.1
Obesobeso	89.5/91.9	80.0/95.8	67.7/93.6	97.1/98.0	-/73.3	73.9/89.2	71.2/ 95.5
Bear	84.2/79.5	81.4/92.2	65.1/90.9	98.8/99.3	-/68.8	76.1/39.2	70.9/ 88.2
Puppy	65.6/73.3	71.4/89.6	55.6/87.2	93.5/98.8	-/43.2	57.4/45.8	61.7/85.8
Mean	78.5/74.7	71.3/90.8	62.2/89.1	90.9/97.8	-/59.3	60.9/58.0	67.6/86.0

Table 5 illustrates our performance on VisA [36] comparing with existing methods. We also achieve 1% improvement for both image and pixel level from the baseline method. Table 6 demonstrates our capability of multi-pose anomaly detection on the challenge MADsim [35]. Without using pose alignment process, we achieve superior performance than other anomaly synthesis based methods.

Qualitative evaluation. We visualize the pixel-wise prediction of our method on MVTecLOCO [4] and MVTecAD [2] in Fig. 6. The reconstruction sub-network generates high quality normal version of the received anomaly input. The localization sub-network precisely reveals the anomaly regions by comparing the difference between the reconstructed images and input. For example, in the second row of the MVTecLOCO results, the anomalous empty bottle is filled with white juice according to the banana sticker. The empty regions excepted the stickers parts are predicted as anomalies. Similarly, the red juice doesn't belong to banana sticker is corrected into white juice and highlighted as anomalies. In the examples of capsule from MVTecAd dataset, the faulty imprint of '500' is recovered both in color image and edge map. The cracked and squeezed capsule are reconstructed into normal appearance. The corresponding anomalous regions are revealed. More results are shown in supplementary material.

4.4. Ablation Study

Table 3 demonstrates the effectiveness of proposed edge manipulation based anomaly synthesis. We carry on ablation experiments on MVTecLOCO with different region selection and edge modification strategies and TPS augmentation. For region selection strategy, we compare the performance of using both or either semantic and arbitrary regions for edge modification. For edge modification strat-

egy, we evaluate the effectiveness of removing, replacing and merging edges in the selected regions. We also evaluate the contribution of TPS warping augmentation on synthetic anomalies. Our baseline is OmniAL [34] that achieves 85.2 image AUROC and 63.2 pixel AUsPRO. By training with the anomalies synthesised by modifying edges in semantic regions, we achieves more than 3% improvement in both image-level detection and pixel-level localization. By combining all edge modification strategies, we improve the performance to 67.0 pixel AUsPRO. The arbitrary region selection strategy increase the diversity of training samples and brings 2.7% improvement. By applying TPS warping augmentation, we achieve 0.7% improvement for image AU-ROC but decrease 2.5% pixel AUsPRO.

5. Conclusion

Different with existing methods, this paper proposes a novel anomaly synthesis method for both unsupervised logical and structural anomaly localization. It generates photo-realistic anomaly images violating underlying logical constraints and fills the gap left by existing structure-oriented anomaly synthesis methods. More specifically, it introduces edge manipulation strategies to balance logical and structural anomaly synthesis. By modifying edges in semantic regions, it easily generates anomalies that break logical constraints, such as missing components. By editing edges in arbitrary regions, it forges varies structural defects. It further improves the anomaly localization performance by introducing edge reconstruction into the network structure. Extensive experiments on the challenge dataset verify the advantages of proposed method.

References

- [1] Kilian Batzner, Lars Heckler, and Rebecca König. Efficient: Accurate visual anomaly detection at millisecond-level latencies. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 128–138, 2024. 6, 7
- [2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec AD - A comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019. Computer Vision Foundation / IEEE. 2, 5, 6, 7, 8
- [3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *CVPR*, pages 4182–4191, 2020. Computer Vision Foundation / IEEE. 3, 6
- [4] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *Int. J. Comput. Vis.*, 130(4):947–969, 2022. 1, 2, 3, 5, 6, 7, 8
- [5] Mircea Cimpoi, Subhansu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *CVPR*, pages 3606–3613, 2014. IEEE Computer Society. 3
- [6] Niv Cohen and Yedid Hoshen. Sub-image anomaly detection with deep pyramid correspondences. *CoRR*, abs/2005.02357. 3, 6
- [7] Niv Cohen, Issar Tzachor, and Yedid Hoshen. Set features for fine-grained anomaly detection. *CoRR*, abs/2302.12245, 2023. 6
- [8] Thomas Defard, Aleksandr Setkov, Angélique Loesch, and Romaric Audigier. Padim: A patch distribution modeling framework for anomaly detection and localization. In *ICPR*, pages 475–489, 2020. Springer. 3, 6
- [9] Giancarlo Di Biase, Hermann Blum, Roland Siegwart, and Cesar Cadena. Pixel-wise anomaly detection in complex driving scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16918–16927, 2021. 3
- [10] Choubo Ding, Guansong Pang, and Chunhua Shen. Catching both gray and black swans: Open-set supervised anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022*, pages 7378–7388. IEEE, 2022. 2
- [11] Gianluca Donato and Serge Belongie. Approximate thin plate spline mappings. In *Computer Vision—ECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part III* 7, pages 21–31. Springer, 2002. 4, 5
- [12] Mariana-Iuliana Georgescu, Radu Tudor Ionescu, Fahad Shahbaz Khan, Marius Popescu, and Mubarak Shah. A background-agnostic framework with adversarial training for abnormal event detection in video. *IEEE Trans. Pattern Anal. Mach. Intell.*, 44(9):4505–4523, 2022. 2
- [13] Denis A. Gudovskiy, Shun Ishizaka, and Kazuki Kozuka. CFLOW-AD: real-time unsupervised anomaly detection with localization via conditional normalizing flows. In *WACV*, pages 1819–1828, 2022. IEEE. 8
- [14] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloé Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross B. Girshick. Segment anything. *CoRR*, abs/2304.02643, 2023. 2, 4, 6
- [15] Chun-Liang Li, Kihyuk Sohn, Jinsung Yoon, and Tomas Pfister. Cutpaste: Self-supervised learning for anomaly detection and localization. In *CVPR*, pages 9664–9674, 2021. Computer Vision Foundation / IEEE. 1, 2, 3, 8
- [16] Tsung-Yi Lin, Priya Goyal, Ross B. Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *CVPR*, pages 2999–3007, 2017. IEEE Computer Society. 5
- [17] Krzysztof Lis, Krishna Nakka, Pascal Fua, and Mathieu Salzmann. Detecting the unexpected via image resynthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 2152–2161, 2019. 3
- [18] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Kai Wang, and Xiang Bai. Richer convolutional features for edge detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3000–3009, 2017. 5
- [19] Prasanta Chandra Mahalanobis. On the generalized distance in statistics. National Institute of Science of India, 1936. 3
- [20] Masoud PourReza, Bahram Mohammadi, Mostafa Khaki, Samir Bouindour, Hichem Snoussi, and Mohammad Sabokrou. G2D: generate to detect anomaly. In *WACV*, pages 2002–2011, 2021. IEEE. 2
- [21] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pages 14318–14328, 2022. 3, 6, 7, 8
- [22] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. *Int. J. Comput. Vis.*, 128(2):336–359, 2020. 3
- [23] Zhuo Su, Wenzhe Liu, Zitong Yu, Dewen Hu, Qing Liao, Qi Tian, Matti Pietikäinen, and Li Liu. Pixel difference networks for efficient edge detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5117–5127, 2021. 4, 5, 7
- [24] Yael Vinker, Eliahu Horwitz, Nir Zabari, and Yedid Hoshen. Image shape manipulation from a single augmented training sample. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 13769–13778, 2021. 4, 5, 7
- [25] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. High-resolution image synthesis and semantic manipulation with conditional gans. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8798–8807, 2018. 4
- [26] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.*, 13(4):600–612, 2004. 5
- [27] Jinjian Wu, Guangming Shi, Weisi Lin, Anmin Liu, and Fei Qi. Just noticeable difference estimation for images with

- free-energy principle. *IEEE Trans. Multim.*, 15(7):1705–1710, 2013. 3, 5
- [28] Yingda Xia, Yi Zhang, Fengze Liu, Wei Shen, and Alan L Yuille. Synthesize then compare: Detecting failures and anomalies for semantic segmentation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part I 16*, pages 145–161. Springer, 2020. 3
- [29] Minghui Yang, Jing Liu, Zhiwei Yang, and Zhaoyang Wu. Slsq: Industrial image anomaly detection by learning better feature embeddings and one-class classification. *arXiv preprint arXiv:2305.00398*, 2023. 3, 6, 7
- [30] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *CoRR*, abs/2206.03687, 2022. 8
- [31] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. Dræm - A discriminatively trained reconstruction embedding for surface anomaly detection. In *ICCV*, pages 8310–8319, 2021. IEEE. 1, 2, 3, 6, 7, 8
- [32] Vitjan Zavrtanik, Matej Kristan, and Danijel Skocaj. DSR - A dual subspace re-projection network for surface anomaly detection. In *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part XXXI*, pages 539–554. Springer, 2022.
- [33] Ying Zhao. Just noticeable learning for unsupervised anomaly localization and detection. In *ICME*, pages In Press, 2022. 2, 3, 5, 6, 7
- [34] Ying Zhao. Omnia: A unified cnn framework for unsupervised anomaly localization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3924–3933, 2023. 1, 2, 3, 5, 6, 7, 8
- [35] Qiang Zhou, Weize Li, Lihan Jiang, Guoliang Wang, Guyue Zhou, Shanghang Zhang, and Hao Zhao. Pad: A dataset and benchmark for pose-agnostic anomaly detection. *arXiv preprint arXiv:2310.07716*, 2023. 5, 6, 7, 8
- [36] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. *CoRR*, abs/2207.14315, 2022. 3, 5, 6, 7, 8