DRÆM – A discriminatively trained reconstruction embedding for surface anomaly detection

Vitjan Zavrtanik Matej Kristan Danijel Skočaj
University of Ljubljana, Faculty of Computer and Information Science
{vitjan.zavrtanik, matej.kristan, danijel.skocaj}@fri.uni-lj.si

Abstract

Visual surface anomaly detection aims to detect local image regions that significantly deviate from normal appearance. Recent surface anomaly detection methods rely on generative models to accurately reconstruct the normal areas and to fail on anomalies. These methods are trained only on anomaly-free images, and often require hand-crafted post-processing steps to localize the anomalies, which prohibits optimizing the feature extraction for maximal detection capability. In addition to reconstructive approach, we cast surface anomaly detection primarily as a discriminative problem and propose a discriminatively trained reconstruction anomaly embedding model (DRÆM). The proposed method learns a joint representation of an anomalous image and its anomaly-free reconstruction, while simultaneously learning a decision boundary between normal and anomalous examples. The method enables direct anomaly localization without the need for additional complicated post-processing of the network output and can be trained using simple and general anomaly simulations. On the challenging MVTec anomaly detection dataset, DRÆM outperforms the current state-of-the-art unsupervised methods by a large margin and even delivers detection performance close to the fully-supervised methods on the widely used DAGM surface-defect detection dataset, while substantially outperforming them in localization accuracy.

1. Introduction

Surface anomaly detection addresses localization of image regions that deviate from a normal appearance (Figure 1). A closely related general anomaly detection problem considers anomalies as *entire images* that significantly differ from the non-anomalous training set images. In contrast, in surface anomaly detection problems, the anomalies occupy only *a small fraction* of image pixels and are typically close to the training set distribution. This is a particularly challenging task, which is common in quality control and surface defect localization applications.

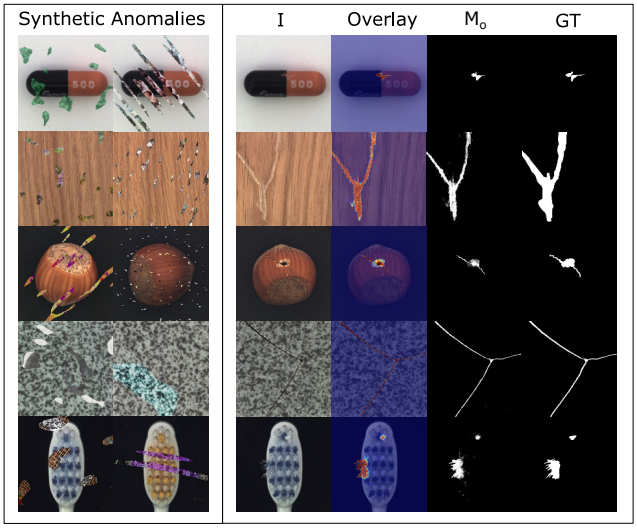

Figure 1. DRÆM estimates the decision boundary between the normal an anomalous pixels solely by training on synthetic anomalies automatically generated on anomaly-free images (left) and generalizes to a variety of real-world anomalies (right). The result (M_o) closely matches the ground truth (GT).

In practice, anomaly appearances may significantly vary, and in applications like quality control, images with anomalies present are rare and manual annotation may be overly time consuming. This leads to highly imbalanced training sets, often containing only anomaly-free images. Significant effort has thus been recently invested in designing robust surface anomaly detection methods that preferably require minimal supervision from manual annotation.

Reconstructive methods, such as Autoencoders [5, 1, 2, 26] and GANs [24, 23], have been extensively explored since they enable learning of a powerful reconstruction subspace, using only anomaly-free images. Relying on poor reconstruction capability of anomalous regions, not observed in training, the anomalies can then be detected by thresholding the difference between the input image and its re-

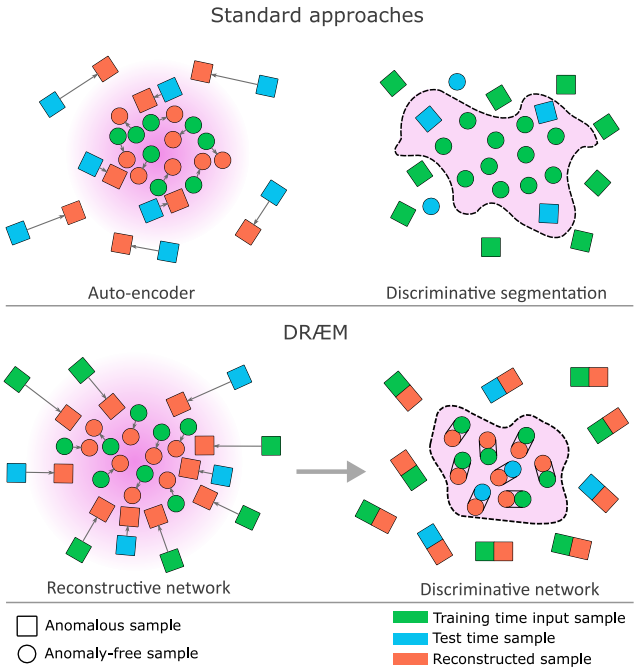

Figure 2. Autoencoders over-generalize to anomalies, while discriminative approaches over-fit to the synthetic anomalies and do not generalize to real data. Our approach jointly discriminatively learns the reconstruction subspace and a hyper-plane over the joint original and reconstructed space using the simulated anomalies and leads to substantially better generalization to real anomalies.

construction. However, determining the presence of anomalies that are not substantially different from normal appearance remains challenging, since these are often well reconstructed, as depicted in Figure 2, top-left.

Recent improvements thus consider the difference between deep features extracted from a general-purpose network and a network specialized for anomaly-free images [4]. Discrimination can also be formulated as a deviation from a dense clustering of non-anomalous textures within the deep subspace [22, 7], as forming such a compact subspace prevents anomalies from being mapped close to anomaly-free samples. A common drawback of the generative methods is that they only learn the model from anomaly-free data, and are not explicitly optimized for discriminative anomaly detection, since positive examples (i.e., anomalies) are not available at training time. Synthetic anomalies could be considered to train discriminative segmentation methods [8, 21], but this leads to over-fitting to synthetic appearances and results in a learned decision boundary that generalizes poorly to real anomalies (Figure 2, top-right).

We hypothesize that over-fitting can be substantially reduced by training a discriminative model over the joint, reconstructed and original, appearance along with the reconstruction subspace. This way the model does not overfit to synthetic appearance, but rather learns a local-appearance-conditioned distance function between the original and reconstructed anomaly appearance, which generalizes well over a range of real anomalies (see Figure 2, bottom).

To validate our hypothesis, we propose, as our main contribution, a new deep surface anomaly detection network, discriminatively trained in an end-to-end manner on synthetically generated just-out-of-distribution patterns, which do not have to faithfully represent the target-domain anomalies. The network is composed of a reconstructive sub-network, followed by a discriminative sub-network (Figure 3). The reconstructive sub-network is trained to learn anomaly-free reconstruction, while the discriminative sub-network learns a discriminative model over the joint appearance of the original and reconstructed images, producing a high-fidelity per-pixel anomaly detection map (Figure 1).

In contrast to related approaches that learn surrogate generative tasks, the proposed model is trained discriminatively, yet does not require the synthetic anomaly appearances to closely match the anomalies at test time and outperforms the recent, more complex, state-of-the-art methods by a large margin.

2. Related work

Many surface anomaly detection methods focus on image reconstruction and detect anomalies based on image reconstruction error [1, 2, 5, 24, 23, 26, 31]. Auto-encoders are commonly used for image reconstruction [5]. In [1, 2, 26] auto-encoders are trained with adversarial losses. The anomaly score of the image is then based on the image reconstruction quality or in the case of adversarially trained auto-encoders, the discriminator output. In [24, 23] a GAN [13] is trained to generate images that fit the training distribution. In [23] an encoder network is additionally trained that finds the latent representation of the input image that minimizes the reconstruction loss when used as the input by the pretrained generator. The anomaly score is then based on the reconstruction quality and the discriminator output. In [29] an interpolation auto-encoder is trained to learn a dense representation space of in-distribution samples. The anomaly score is then based on a discriminator, trained to estimate the distance between the input-input and input-output joint distributions, however the approach to surface anomaly detection remains generative as the discriminator evaluates the reconstruction quality.

Instead of the commonly used image space reconstruction, the reconstruction of pretrained network features can also be used for surface anomaly detection [4, 25]. Anomalies are detected based on the assumption that features of a pre-trained network will not be faithfully reconstructued by another network trained only on anomaly-free images. Alternatively [20, 11] propose surface anomaly detection as identifying significant deviations from a Gaussian fitted to

anomaly-free features of a pre-trained network. This requires a unimodal distribution of the anomaly-free visual features which is problematic on diverse datasets. [16] propose a one-class variational auto-encoder gradient-based attention maps as output anomaly maps. However the method is sensitive to subtle anomalies close to the normal sample distribution.

Recently Patch-based one-class classification methods have been considered for surface anomaly detection [30]. These are based on one-class methods [22, 7] which attempt to estimate a decision boundary around anomaly-free data that separates it from anomalous samples by assuming a unimodal distribution of the anomaly-free data. This assumption is often violated in surface anomaly data.

3. DRÆM

The proposed discriminative joint reconstruction-anomaly embedding method (DRÆM) is composed from a reconstructive and a discriminative sub-networks (see Figure 3). The reconstructive sub-network is trained to implicitly detect and reconstruct the anomalies with semantically plausible anomaly-free content, while keeping the non-anomalous regions of the input image unchanged. Simultaneously, the discriminative sub-network learns a joint reconstruction-anomaly embedding and produces accurate anomaly segmentation maps from the concatenated reconstructed and original appearance. Anomalous training examples are created by a conceptually simple process that simulates anomalies on anomaly-free images. This anomaly generation method provides an arbitrary amount of anomalous samples as well as pixel-perfect anomaly segmentation maps which can be used for training the proposed method without real anomalous samples.

3.1. Reconstructive sub-network

The reconstructive sub-network is formulated as an encoder-decoder architecture that converts the local patterns of an input image into patterns closer to the distribution of normal samples. The network is trained to reconstruct the original image I from an artificially corrupted version I_a obtained by a simulator (see Section 3.3).

An l_2 loss is often used in reconstruction based anomaly detection methods [1, 2], however this assumes an independence between neighboring pixels, therefore a patch based SSIM [27] loss is additionally used as in [5, 31]:

$$L_{SSIM}(I, I_r) = \frac{1}{N_p} \sum_{i=1}^{H} \sum_{j=1}^{W} 1 - SSIM\big(I, I_r\big)_{(i,j)}, \quad (1)$$

where H and W are the height and width of image I, respectively. N_p is equal to the number of pixels in I. I_r is the reconstructed image output by the network. $SSIM(I, I_r)_{(i,j)}$ is the SSIM value for patches of I and I_r, centered at image coordinates (i, j). The reconstruction loss is therefore:

$$L_{rec}(I, I_r) = \lambda L_{SSIM}(I, I_r) + l_2(I, I_r), \quad (2)$$

where λ is a loss balancing hyper-parameter.

Note that an additional training signal is acquired from the downstream discriminative network (Section 3.2), which performs anomaly localization by detecting the reconstruction difference.

3.2. Discriminative sub-network

The discriminative sub-network uses U-Net [21]-like architecture. The sub-network input I_c is defined as the channel-wise concatenation of the reconstructive sub-network output I_r and the input image I. Due to the normality-restoring property of the reconstructive sub-network, the joint appearance of I and I_r differs significantly in anomalous images, providing the information necessary for anomaly segmentation. In reconstruction-based anomaly detection methods anomaly maps are obtained using similarity functions such as SSIM [27] to compare the original image to its reconstruction, however a surface anomaly detection-specific similarity measure is difficult to hand-craft. In contrast, the discriminative sub-network learns the appropriate distance measure automatically. The network outputs an anomaly score map M_o of the same size as I. Focal Loss [14] (L_{seg}) is applied on the discriminative sub-network output to increase robustness towards accurate segmentation of hard examples.

Considering both the segmentation and the reconstructive objectives of the two sub-networks, the total loss used in training DRÆM is

$$L(I, I_r, M_a, M) = L_{rec}(I, I_r) + L_{seg}(M_a, M), \quad (3)$$

where M_a and M are the ground truth and the output anomaly segmentation masks, respectively.

3.3. Simulated anomaly generation

DRÆM does not require simulations to realistically reflect the real anomaly appearance in the target domain, but rather to generate just-out-of-distribution appearances, which allow learning the appropriate *distance function* to recognize the anomaly by its deviation from normality. The proposed anomaly simulator follows this paradigm.

A noise image is generated by a Perlin noise generator [18] to capture a variety of anomaly shapes (Figure 4, P) and binarized by a threshod sampled uniformly at random (Figure 4, M_a) into an anomaly map M_a. The anomaly texture source image A is sampled from an anomaly source image dataset which is unrelated to the input image distribution (Figure 4, A). Random augmentation sampling, inspired by RandAugment [10], is then applied by a set

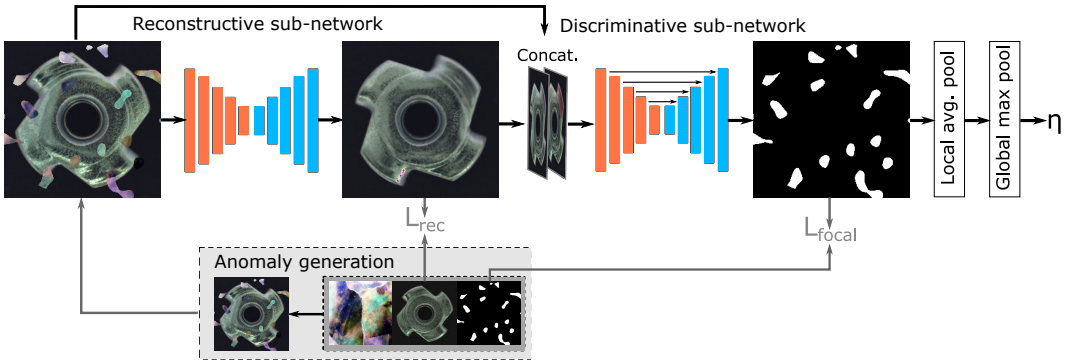

Figure 3. The anomaly detection process of the proposed method. First anomalous regions are implicitly detected and inpainted by the reconstructive sub-network trained using L_{rec}. The output of the reconstructive sub-network and the input image are then concatenated and fed into the discriminative sub-network. The segmentation network, trained using the Focal loss L_{focal}[14], localizes the anomalous region and produces an anomaly map. The image level anomaly score η is acquired from the anomaly score map.

of 3 random augmentation functions sampled from the set: {*posterize, sharpness, solarize, equalize, brightness change, color change, auto-contrast*}. The augmented texture image A is masked with the anomaly map M_a and blended with I to create anomalies that are just-out-of-distribution, and thus help tighten the decision boundary in the trained network. The augmented training image I_a is therefore defined as

$$I_a = \overline{M}_a \odot I + (1-\beta)(M_a \odot I) + \beta(M_a \odot A), \quad (4)$$

where $\overline{M}_a$ is the inverse of M_a, $\odot$ is the element-wise multiplication operation and β is the opacity parameter in blending. This parameter is sampled uniformly from an interval, i.e., $\beta \in [0.1, 1.0]$. The randomized blending and augmentation afford generating diverse anomalous images from as little as *a single texture* (see Figure 5).

The above described simulator thus generates training sample triplets containing the original anomaly-free image I, the augmented image containing simulated anomalies I_a and the pixel-perfect anomaly mask M_a.

3.4. Surface anomaly localization and detection

The output of the discriminative sub-network is a pixel-level anomaly detection mask M_o, which can be interpreted in a straight-forward way for the image-level anomaly score estimation, i.e., whether an anomaly is present in the image.

First, M_o is smoothed by a mean filter convolution layer to aggregate the local anomaly response information. The final image-level anomaly score η is computed by taking the maximum value of the smoothed anomaly score map:

$$\eta = max\big(M_o * f_{s_f \times s_f}\big) \quad , \quad (5)$$

where $f_{s_f \times s_f}$ is a mean filter of size $s_f \times s_f$ and $*$ is the convolution operator. In a preliminary study, we trained a classification network for the image-level anomaly classification, but did not observe improvements over the direct score estimation method (5).

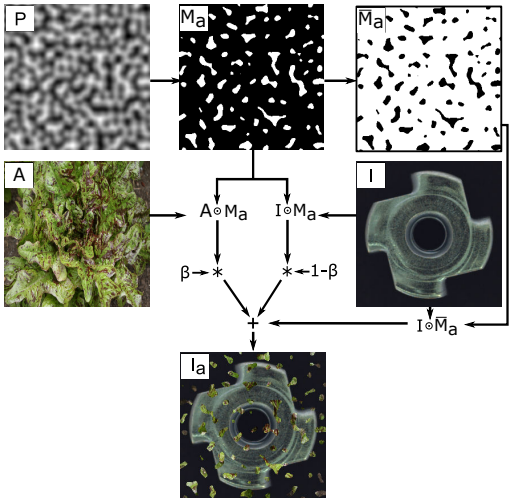

Figure 4. Simulated anomaly generation process. The binary anomaly mask M_a is generated from Perlin noise P. The anomalous regions are sampled from A according to M_a and placed on the anomaly free image I to generate the anomalous image I_a.

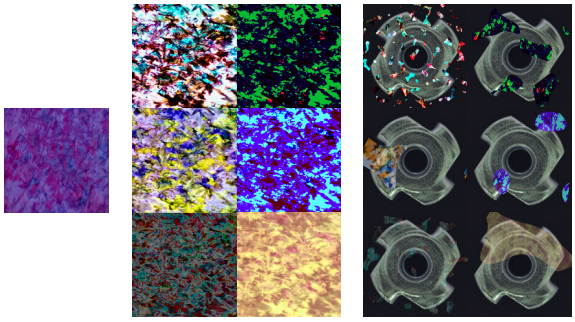

Figure 5. The original anomaly source image (left) can be augmented several times (center) to generate a wide variety of simulated anomalous regions (right).

4. Experiments

DRÆM is extensively evaluated and compared with the recent state-of-the-art on unsupervised surface anomaly detection and localization. Additionally, individual components of the proposed method and the effectiveness of training on simulated anomalies are evaluated by an ablation study. Finally, the results are placed in a broader perspective by comparing DRÆM with state-of-the-art weakly-supervised and fully-supervised surface-defect detection methods.

4.1. Comparison with unsupervised methods

DRÆM is evaluated on the recent challenging MVTec anomaly detection dataset [3], which has been established as a standard benchmark dataset for evaluating unsupervised surface anomaly detection methods. We evaluate DRÆM on the tasks of surface anomaly detection and localisation. The MVTec dataset contains 15 object classes with a diverse set anomalies which enables a general evaluation of surface anomaly detection methods. Anomalous examples of the MVTec dataset are shown in Figure 8. For evaluation, the standard metric in anomaly detection, AUROC, is used. Image-level AUROC is used for anomaly detection and a pixel-based AUROC for evaluating anomaly localization [5, 24, 17, 26]. The AUROC, however, does not reflect the localization accuracy well in surface anomaly detection setups, where only a small fraction of pixels are anomalous. The reason is that false positive rate is dominated by the a-priori very high number of non-anomalous pixels and is thus kept low despite of false positive detections. We thus additionally report the pixel-wise average precision metric (AP), which is more appropriate for highly imbalanced classes and in particular for surface anomaly detection, where the precision plays an important role.

In our experiments, the network is trained for 700 epochs on the MVTec anomaly detection dataset [3]. The learning rate is set to 10^{-4} and is multiplied by 0.1 after 400 and 600 epochs. Image rotation in the range of $(-45, 45)$ degrees is used as a data augmentation method on anomaly free images during training to alleviate overfitting due to the relatively small anomaly-free training set size. The Describable Textures Dataset [9] is used as the anomaly source dataset.

A number of obtained qualitative examples are presented in Figure 8. As one can observe, the obtained anomaly masks are very detailed and resemble the given ground truth labels to a high degree of accuracy. Consequently, DRÆM achieves state-of-the-art quantitative results across all MVTec classes for surface anomaly detection as well as localization.

Surface Anomaly Detection. Table 1 quantitatively compares DRÆM with recent approaches on the task of *image-level surface anomaly detection*. DRÆM significantly outperforms all recent surface anomaly detection methods, achieving the highest AUROC in 9 out of 15 classes and achieving comparable results in the other classes. It surpasses the previous best state-of-the-art approach by 2.5 percentage points. The reduced performance in some classes could be explained by particularly difficult anomalies that are close to the normal image distribution. The absence of a part of the object is especially difficult to detect. Regions, where the object features are missing, usually contain other commonly occurring features. This makes such anomalies difficult to distinguish from anomaly-free regions. An example of this can be seen in Figure 6, where some of the transistor leads had been cut. The ground truth marks the area where the broken lead should be as anomalous. DRÆM only detects anomalous features in a small region of the cut lead, as the background features are common during training.

Class	[1]	[26]	[4]	[31]	[20]	[11]	DRÆM
bottle	79.4	98.3	99.0	99.9	**100**	99.9	99.2
capsule	72.1	68.7	86.1	88.4	92.3	91.3	**98.5**
grid	74.3	86.7	81.0	99.6	92.9	96.7	**99.9**
leather	80.8	94.4	88.2	**100**	**100**	**100**	**100**
pill	67.1	76.8	87.9	83.8	83.4	93.3	**98.9**
tile	72.0	96.1	99.1	98.7	97.4	98.1	**99.6**
transistor	80.8	79.4	81.8	90.9	95.9	**97.4**	93.1
zipper	74.4	78.1	91.9	98.1	97.9	90.3	**100**
cable	71.1	66.5	86.2	81.9	**94.0**	92.7	91.8
carpet	82.1	90.3	91.6	84.2	95.5	**99.8**	97.0
hazelnut	87.4	100	93.1	83.3	98.7	92.0	**100.0**
metal nut	69.4	81.5	82.0	88.5	93.1	**98.7**	**98.7**
screw	**100**	**100**	54.9	84.5	81.2	85.8	93.9
toothbrush	70.0	95.0	95.3	**100**	95.8	96.1	**100**
wood	92.0	97.9	97.7	93.0	97.6	**99.2**	99.1
avg	78.2	87.3	87.7	91.7	94.4	95.5	**98.0**

Table 1. Results for the task of surface anomaly detection on the MVTec dataset (AUROC). An average score over all classes is also reported the last row (avg).

Anomaly Localization. Table 2 compares DRÆM to the recent state-of-the-art on the task of *pixel-level surface anomaly detection*. DRÆM achieves comparable results to the previous best-performing methods in terms of AUROC scores and surpasses the state-of-the-art by 13.4 percentage points in terms of AP. A better AP score is achieved in 11 out of 15 classes and is comparable to the state-of-the-art in other classes. A qualitative comparison with the state-of-the-art method Uninformed Students [4] and PaDim [11] is shown in Figure 7. DRÆM achieves a significant improvement in anomaly segmentation accuracy.

A detailed inspection showed that some of the detection errors can be attributed to the inaccurate ground truth labels on ambiguous anomalies. An example of this is shown in Figure 6, where the ground truth covers the entire surface of the pill, yet only the yellow dots are anomalous. DRÆM produces an anomaly map that correctly localizes the yellow dots, but the discrepancy with the ground truth mask

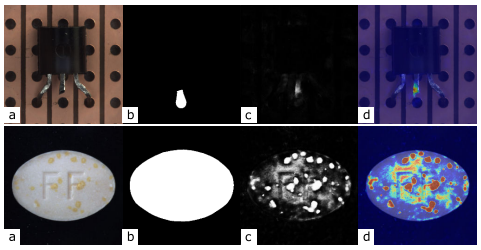

Figure 6. The original image (a) contains anomalies which are difficult to mark in the ground truth mask (b) which causes a discrepancy between the ground truth and the output anomaly map (c,d).

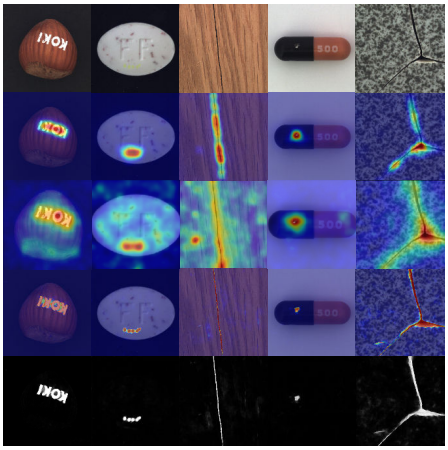

Figure 7. The anomalous images are shown in the first row. The middle three rows show the anomaly maps generated by our implementation of Uninformed Students [4], PaDim [11] and DRÆM, respectively. The last row shows the direct anomaly map output of DRÆM.

increases the performance error. These annotation ambiguities also impact the AP score of the evaluated methods.

Class	US[4]	RIAD[31]	PaDim[11]	DRÆM
bottle	97.8 / 74.2	98.4 / 76.4	98.2 / 77.3	**99.1** / **86.5**
capsule	96.8 / 25.9	92.8 / 38.2	**98.6** / 46.7	94.3 / **49.4**
grid	89.9 / 10.1	98.8 / 36.4	97.1 / 35.7	**99.7** / **65.7**
leather	97.8 / 40.9	99.4 / 49.1	**99.0** / 53.5	98.6 / **75.3**
pill	96.5 / **62.0**	95.7 / 51.6	95.7 / 61.2	**97.6** / 48.5
tile	92.5 / 65.3	89.1 / 52.6	94.1 / 52.4	**99.2** / **92.3**
transistor	73.7 / 27.1	87.7 / 39.2	**97.6** / **72.0**	90.9 / 50.7
zipper	95.6 / 36.1	97.8 / 63.4	98.4 / 58.2	**98.8** / **81.5**
cable	91.9 / 48.2	84.2 / 24.4	**96.7** / 45.4	94.7 / **52.4**
carpet	93.5 / 52.2	96.3 / **61.4**	**99.0** / 60.7	95.5 / 53.5
hazelnut	98.2 / 57.8	96.1 / 33.8	98.1 / 61.1	**99.7** / **92.9**
metal nut	97.2 / 83.5	92.5 / 64.3	97.3 / 77.4	**99.5** / **96.3**
screw	97.4 / 7.8	**98.8** / 43.9	98.4 / 21.7	97.6 / **58.2**
toothbrush	97.9 / 37.7	**98.9** / 50.6	98.8 / **54.7**	98.1 / 44.7
wood	92.1 / 53.3	85.8 / 38.2	94.1 / 46.3	**96.4** / **77.7**
avg	93.9 / 45.5	94.2 / 48.2	**97.4** / 55.0	97.3 / **68.4**

Table 2. Results for the task of anomaly localization on the MVTec dataset (AUROC / AP).

4.2. Ablation Study

The DRÆM design choices are analyzed by groups of experiments evaluating (i) the method architecture, (ii) the choice of anomaly appearance patterns and (iii) low perturbation example generation. Results are visually grouped by shades of gray in Table 3.

Architecture. The DRÆM reconstructive sub-network impact on the downstream surface anomaly detection performance is evaluated by removing it from the pipeline and training the discriminative sub-network alone. The results are shown in Table 3, experiment Disc. Note a reduction in performance in comparison to the full DRÆM architecture (Table 3, experiment DRÆM). The performance drop is due to overfitting of the discriminative sub-network to the simulated anomalies, which are not a faithful representation of the real ones.

Next, the discriminative power of the reconstructive sub-network alone is analyzed by evaluating it as an auto-encoder-based surface anomaly detector. The reconstructed image output of the sub-network is compared to the input image using the SSIM function [27] to generate the anomaly map. The results of this approach are shown in Table 3, experiment Recon.-AE. Recon.-AE outperforms the recent auto-encoder-based surface anomaly detection method AE-SSIM[5] (see results in Table 2) This suggests that simulated anomaly training introduces additional information into the auto-encoder-based training, but judging by the performance gap to DRÆM, the SSIM similarity function may not be optimal for extraction of the anomaly information. Indeed, using the recently proposed similarity function MS-GMS [31] (Recon.-AE$_{MSGMS}$) improves the performance, but the results are still significantly worse than when using the entire DRÆM architecture, which indicates that both reconstructive and discriminative parts are required for optimal results.

To further emphasize the contribution of the DRÆM backbone, we replace it entirely by the recent state-of-the-art supervised discriminative surface anomaly detection network [6] and re-train with the simulated anomalies (Table 3, Božič *et al.*). Performance substantially drops, which further supports the power of learning the *anomaly deviation extent* from normality rather than the anomaly or normality *appearance*.

Anomaly appearance patterns. DRÆM is re-trained using ImageNet [12] as the texture source in the anomaly simulator to study the influence of the anomaly generation dataset (DRÆM$_{ImageNet}$ in Table 3). Results are comparable to using the much smaller DTD [9] dataset. Figure 9 shows the performance at various anomaly source dataset sizes. Results suggest that the augmentation and opacity randomization substantially contribute to performance allowing remarkably small number of texture images (less than 10). As an extreme case, the anomaly textures are

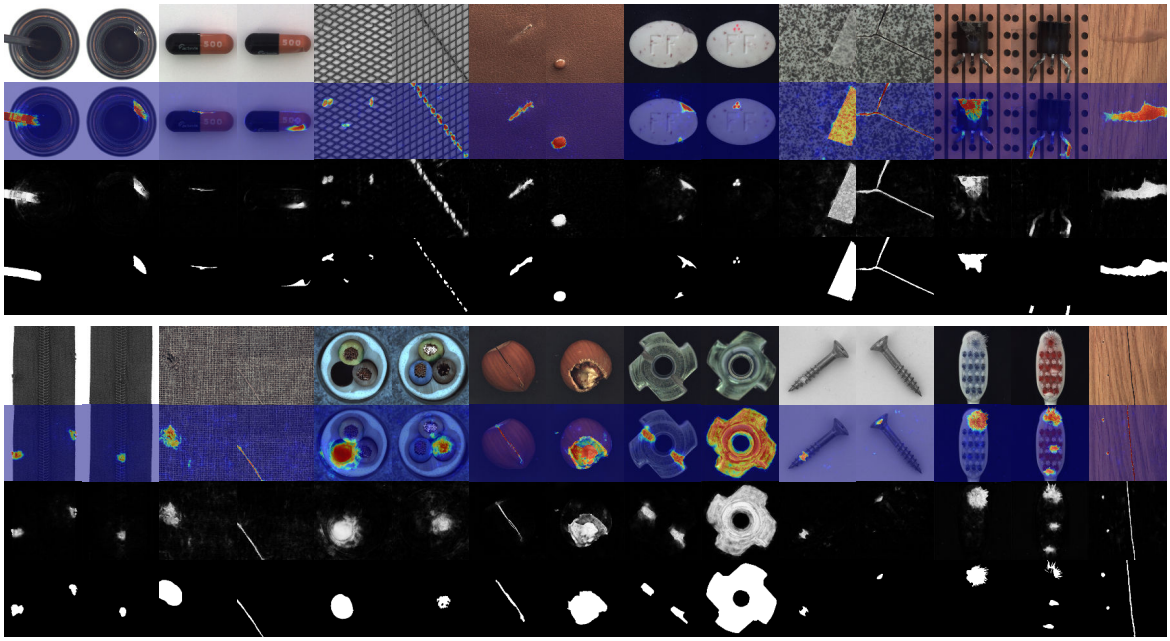

Figure 8. Qualitative examples. The original image, the anomaly map overlay, the anomaly map and the ground truth map are shown.

	Architecture			Anomaly Generation					Results	
Method	Recon. Net.	Discr. Net.	Augmentation	β	ImageNet	DTD	Perlin	Rectangle	Det.	Loc.
Disc.		✓	✓	✓		✓	✓		93.9	92.7 / 62.5
Recon.-AE	✓		✓	✓		✓	✓		83.9	89.7 / 47.5
Recon.-AE$_{MSGMS}$	✓		✓	✓		✓	✓		90.7	93.4 / 50.9
Božič *et al.* [6]			✓	✓		✓	✓		92.8	93.9 / 60.7
DRÆM$_{ImageNet}$	✓	✓	✓	✓	✓		✓		97.9	97.0 / 67.9
DRÆM$_{color}$	✓	✓		✓			✓		96.2	92.6 / 56.5
DRÆM$_{rect}$	✓	✓	✓	✓		✓		✓	96.9	96.8 / 65.1
DRÆM$_{no_aug}$	✓	✓				✓	✓		97.4	94.5 / 64.3
DRÆM$_{img_aug}$	✓	✓	✓			✓	✓		97.4	95.0 / 64.5
DRÆM$_{\beta}$	✓	✓		✓		✓	✓		97.9	97.1 / **68.4**
DRÆM	✓	✓	✓	✓		✓	✓		**98.0**	**97.3 / 68.4**

Table 3. Surface anomaly detection (Det.) and localization (Loc.) experiments of the ablation study grouped by shades of gray into (i) method architecture, (ii) anomaly source dataset, (iii) hard simulated anomaly generation, (iv) simulated anomaly shape, and (v) the performance of DRÆM for reference.

generated as homogeneous regions of a randomly sampled color (DRÆM$_{color}$). Note that DRÆM$_{color}$ still achieves state-of-the-art results, further suggesting that DRÆM does not require simulations to closely match the real anomalies.

The impact of the anomaly shape generator is evaluated by replacing the Perlin noise generator by a rectangular region generator. The anomaly mask is thus generated by sampling multiple rectangular areas for the anomalous regions (DRÆM$_{rect}$ in Table 3). Training on rectangular anomalies causes only a slight performance drop and suggests that the simulated anomaly shape does not have to be realistic to generalize well to real world anomalies. Examples of anomalies generated in anomaly appearance ablation experiments are shown in Figure 10.

Low perturbation examples. The anomaly source image augmentation and the opacity randomization are responsible for tightening the decision boundary around the anomaly-free training distribution. Table 3 reports the results of DRÆM variants trained (i) without image augmentation and opacity randomization (DRÆM$_{no_aug}$), (ii) using only image augmentation (DRÆM$_{img_aug}$) and (iii) using only opacity randomization (DRÆM$_{\beta}$). There is a significant localization performance gap between DRÆM$_{no_aug}$ and DRÆM, however, this can be significantly narrowed by using the opacity randomization in training even without image data augmentation.

4.3. Comparison with supervised methods

Supervised methods require anomaly annotations at training time and cannot be evaluated on MVTec. We thus compare DRÆM with the supervised methods on the DAGM dataset [28] that contains 10 textured object classes

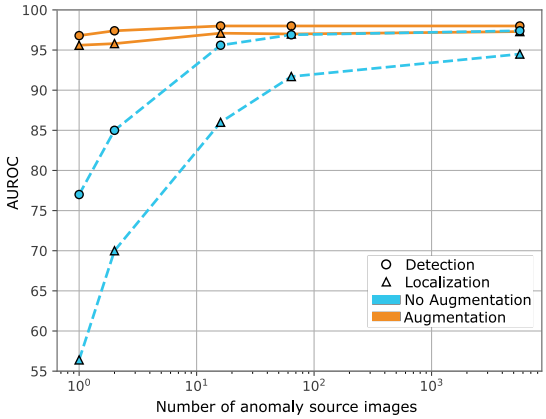

Figure 9. DRÆM achieves a remarkable detection and localization performance already at as low as 10 texture source images in the simulator when augmentation is applied.

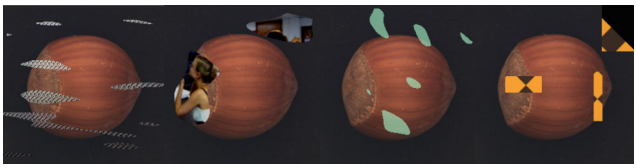

Figure 10. Anomalies simulated using the DTD [9] (DRÆM), ImageNet [12] (DRÆM$_{ImageNet}$), homogeneous color regions (DRÆM$_{color}$) and rectangular masks (DRÆM$_{rect}$), from left to right.

with small anomalies visually very similar to the background, which makes the dataset particularly challenging for the unsupervised methods.

DRÆM is trained only on anomaly-free training samples using the same parameters as in previous experiments. The standard evaluation protocol on this dataset [19, 32, 15, 6] is used – the challenge is to classify whether the image contains the anomaly; localization accuracy is not measured, since the anomalies are only coarsely labeled.

Table 4 shows that the best fully supervised methods nearly perfectly classify anomalous images, while the state-of-the-art unsupervised methods like RIAD [31] and US [4] struggle with subtle anomalies on highly textured regions[1]. DRÆM significantly outperforms these methods, and even the weakly supervised CADN [32] by a large margin, obtaining classification performance close to the best fully-supervised methods, which is a remarkable result.

Furthermore, DRÆM by far outperforms all supervised methods in terms of anomaly localization accuracy on this dataset. Since the training images are only coarsely annotated with ellipses that approximately cover the surface defects and contain background, the supervised methods produce inaccurate localization in test images as well. In contrast, DRÆM does not use the labels at all, and thus produces significantly more accurate anomaly maps, as shown in Figure 11.

[1]Please see the supplementary material for additional qualitative results.

	Methods	AUROC	TPR	TNR	CA
Unsup.	RIAD [31]	78.6	79.2	69.1	70.4
	US [4]	72.5	72.6	65.3	66.2
	MAD [20]	82.4	78.7	85.7	66.2
	PaDim [11]	95.0	83.3	97.5	95.7
	DRÆM	**99.0**	**96.5**	**99.4**	**98.5**
Sup.	CADN [32]	-	-	-	89.1
	Rački *et al.* [19]	99.6	99.9	99.5	-
	Lin *et al.* [15]	99.0	99.4	99.9	-
	Božič *et al.* [6]	**100**	**100**	**100**	**100**

Table 4. DRÆM outperforms unsupervised methods on DAGM dataset and performs on par with fully supervised ones.

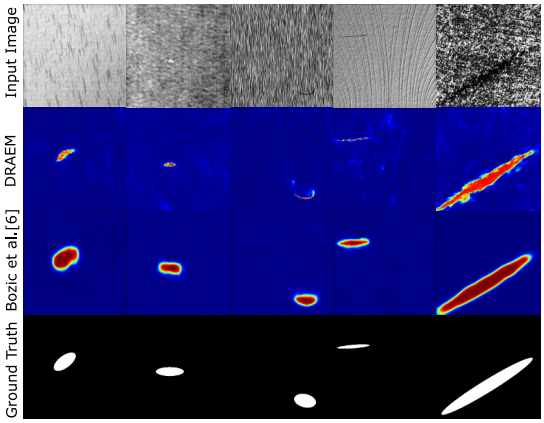

Figure 11. Supervised methods replicate the approximate ground truth training annotations, leading to a low localization accuracy. DRÆM does not use the ground truth, yet produces far better localization.

5. Conclusion

A discriminative end-to-end trainable surface anomaly detection and localization method DRÆM was presented. DRÆM outperforms the current state-of-the-art on the MVTec dataset [3] by 2.5 AUROC points on the surface anomaly detection task and by 13.5 AP points on the localization task. On the DAGM dataset [28], DRÆM delivers anomalous image classification accuracy close to fully supervised methods, while outperforming them in localization accuracy. This is a remarkable result since DRÆM is not trained on real anomalies. In fact, a detailed analysis shows that our paradigm of learning a joint reconstruction-anomaly embedding through a reconstructive sub-network significantly improves the results over standard methods and that an accurate decision boundary can be well estimated by learning the extent of deviation from reconstruction on simple simulations rather than learning either the normality or real anomaly appearance.

References

[1] Samet Akcay, Amir Atapour-Abarghouei, and Toby P Breckon. GANomaly: Semi-supervised anomaly detection via adversarial training. In *Asian Conference on Computer Vision*, pages 622–637. Springer, 2018. 1, 2, 3, 5

[2] Samet Akçay, Amir Atapour-Abarghouei, and Toby P Breckon. Skip-GANomaly: Skip connected and adversarially trained encoder-decoder anomaly detection. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, jul 2019. 1, 2, 3

[3] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. MVTec AD – A Comprehensive Real-World Dataset for Unsupervised Anomaly Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9592–9600, 2019. 5, 8

[4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020. 2, 5, 6, 8

[5] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *14th International Joint Conference on Computer Vision, Imaging and Computer Graphics theory and Applications*, volume 5, pages 372–380, 2018. 1, 2, 3, 5, 6

[6] Jakob Božič, Domen Tabernik, and Danijel Skočaj. End-to-end training of a two-stage neural network for defect detection. *25th International Conference on Pattern Recognition ICPR*, 2020. 6, 7, 8

[7] Raghavendra Chalapathy, Aditya Krishna Menon, and Sanjay Chawla. Anomaly detection using one-class neural networks. *arXiv preprint arXiv:1802.06360*, 2018. 2, 3

[8] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *arXiv preprint arXiv:1706.05587*, 2017. 2

[9] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, and Andrea Vedaldi. Describing textures in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3606–3613, 2014. 5, 6, 8

[10] Ekin D Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V Le. Randaugment: Practical automated data augmentation with a reduced search space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 702–703, 2020. 3

[11] Thomas Defard, Aleksandr Setkov, Angelique Loesch, and Romaric Audigier. Padim: a patch distribution modeling framework for anomaly detection and localization. In *1st International Workshop on Industrial Machine Learning, ICPR 2020*, 2020. 2, 5, 6, 8

[12] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 6, 8

[13] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014. 2

[14] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017. 3, 4

[15] Zesheng Lin, Hongxia Ye, Bin Zhan, and Xiaofeng Huang. An efficient network for surface defect detection. *Applied Sciences*, 10(17):6085, 2020. 8

[16] Wenqian Liu, Runze Li, Meng Zheng, Srikrishna Karanam, Ziyan Wu, Bir Bhanu, Richard J Radke, and Octavia Camps. Towards visually explaining variational autoencoders. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8642–8651, 2020. 3

[17] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6536–6545, 2018. 5

[18] Ken Perlin. An image synthesizer. *ACM Siggraph Computer Graphics*, 19(3):287–296, 1985. 3

[19] Domen Rački, Dejan Tomaževič, and Danijel Skočaj. A compact convolutional neural network for textured surface anomaly detection. In *2018 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1331–1339, 2018. 8

[20] Oliver Rippel, Patrick Mertens, and Dorit Merhof. Modeling the distribution of normal data in pre-trained deep features for anomaly detection. *ICPR*, 2020. 2, 5, 8

[21] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015. 2, 3

[22] Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pages 4393–4402, 2018. 2, 3

[23] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical image analysis*, 54:30–44, 2019. 1, 2

[24] Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Ursula Schmidt-Erfurth, and Georg Langs. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In *International Conference on Information Processing in Medical Imaging*, pages 146–157. Springer, 2017. 1, 2, 5

[25] Yong Shi, Jie Yang, and Zhiquan Qi. Unsupervised anomaly segmentation via deep feature reconstruction. *Neurocomputing*, 424:9–22, 2021. 2

[26] Ta-Wei Tang, Wei-Han Kuo, Chien-Fang Lan, Jauh-Hsiang nad Ding, Hakiem Hsu, and Hong-Tsu Young. Anomaly detection neural network with dual auto-encoders gan and its

industrial inspection applications. *Sensors*, 20(12), 2020. 1, 2, 5

[27] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004. 3, 6

[28] M Wieler and T Hahn. Weakly supervised learning for industrial optical inspection, 2007. 7, 8

[29] Yexin Wu, Yogesh Balaji, Bhanukiran Vinzamuri, and Soheil Feizi. Mirrored autoencoders with simplex interpolation for unsupervised anomaly detection. *Proceedings of the European Conference on Computer Vision (ECCV)*, 2020. 2

[30] Jihun Yi and Sungroh Yoon. Patch svdd: Patch-level svdd for anomaly detection and segmentation. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 3

[31] Vitjan Zavrtanik, Matej Kristan, and Danijel Skočaj. Reconstruction by inpainting for visual anomaly detection. *Pattern Recognition*, 2020. 2, 3, 5, 6, 8

[32] Jiabin Zhang, Hu Su, Wei Zou, Xinyi Gong, Zhengtao Zhang, and Fei Shen. Cadn: A weakly supervised learning-based category-aware object detection network for surface defect detection. *Pattern Recognition*, 109:107571, 2021. 8