Paper Review

DRÆM – A discriminatively trained reconstruction embedding for surface

YeongHyeon Park

Department of Electrical and Computer Engineering

SungKyunKwan University

Summary

Zavrtanik, Vitjan, Matej Kristan, and Danijel Skočaj. "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection." Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV). 2021.

DRÆM

2021 ICCV OCTOBER 11-17 VIRTUAL

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

RIAD

Reconstruction by inpainting for visual anomaly detection

Vitjan Zavrtanik, Matej Kristan, Danijel Skočaj

Faculty of Computer and Information Science, University of Ljubljana, Večna pot 113, Ljubljana 1000, Slovenia

A R T I C L E I N F O

Article history:
Received 29 May 2020
Revised 22 September 2020
Accepted 14 October 2020
Available online 17 October 2020

Keywords:
Anomaly detection
Video anomaly detection
Inpainting
CNN

A B S T R A C T

Visual anomaly detection addresses the problem of classification or localization of regions in an image that deviate from their normal appearance. A popular approach trains an auto-encoder on anomaly-free images and performs anomaly detection by calculating the difference between the input and the reconstructed image. This approach assumes that the auto-encoder will be unable to accurately reconstruct anomalous regions. But in practice neural networks generalize well even to anomalies and reconstruct them sufficiently well, thus reducing the detection capabilities. Accurate reconstruction is far less likely if the anomaly pixels were not visible to the auto-encoder. We thus cast anomaly detection as a self-supervised reconstruction-by-inpainting problem. Our approach (RIAD) randomly removes partial image regions and reconstructs the image from partial inpaintings, thus addressing the drawbacks of auto-enocoding methods. RIAD is extensively evaluated on several benchmarks and sets a new state-of-the art on a recent highly challenging anomaly detection benchmark.

About Vitjan Zavrtanik

Vitjan Zavrtanik

University of Ljubljana
Verified email at fri.uni-lj.si
Computer vision anomaly detection

TITLE	CITED BY	YEAR	
A Low-Shot Object Counting Network With Iterative Prototype Adaptation N Djukic, A Lukezic, V Zavrtanik, M Kristan arXiv preprint arXiv:2211.08217		2022	
DSR–A dual subspace re-projection network for surface anomaly detection V Zavrtanik, M Kristan, D Skočaj Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel …	9	2022	DSR, ECCV, Oct. 2022
Dsr—A dual subspace re-projection network for surface anomaly detection supplementary material V Zavrtanik, M Kristan, D Skocaj	1	2022	
Reconstruction by inpainting for visual anomaly detection V Zavrtanik, M Kristan, D Skočaj Pattern Recognition 112, 107706	174	2021	RIAD, Pattern Recognition Oct. 2020
Draem-a discriminatively trained reconstruction embedding for surface anomaly detection V Zavrtanik, M Kristan, D Skočaj Proceedings of the IEEE/CVF International Conference on Computer Vision …	117	2021	DRAEM, ICCV, Oct. 2021
A segmentation-based approach for polyp counting in the wild V Zavrtanik, M Vodopivec, M Kristan Engineering Applications of Artificial Intelligence 88, 103399	6	2020	
Segmentacijska konvolucijska nevronska mreža za štetje polipov na slikah V Zavrtanik Univerza v Ljubljani		2018	
Segmentacijska konvolucijska nevronska mreža za štetje polipov na slikah: magistrsko delo: magistrski program druge stopnje Računalništvo in informatika V Zavrtanik V. Zavrtanik		2018	MS
Učinkovito generiranje hipotetičnih slikovnih regij za detekcijo polipov V Zavrtanik Univerza v Ljubljani		2016	
Učinkovito generiranje hipotetičnih slikovnih regij za detekcijo polipov: diplomsko delo: univerzitetni študijski program prve stopnje Računalništvo in informatika V Zavrtanik V. Zavrtanik		2016	BS

Concept

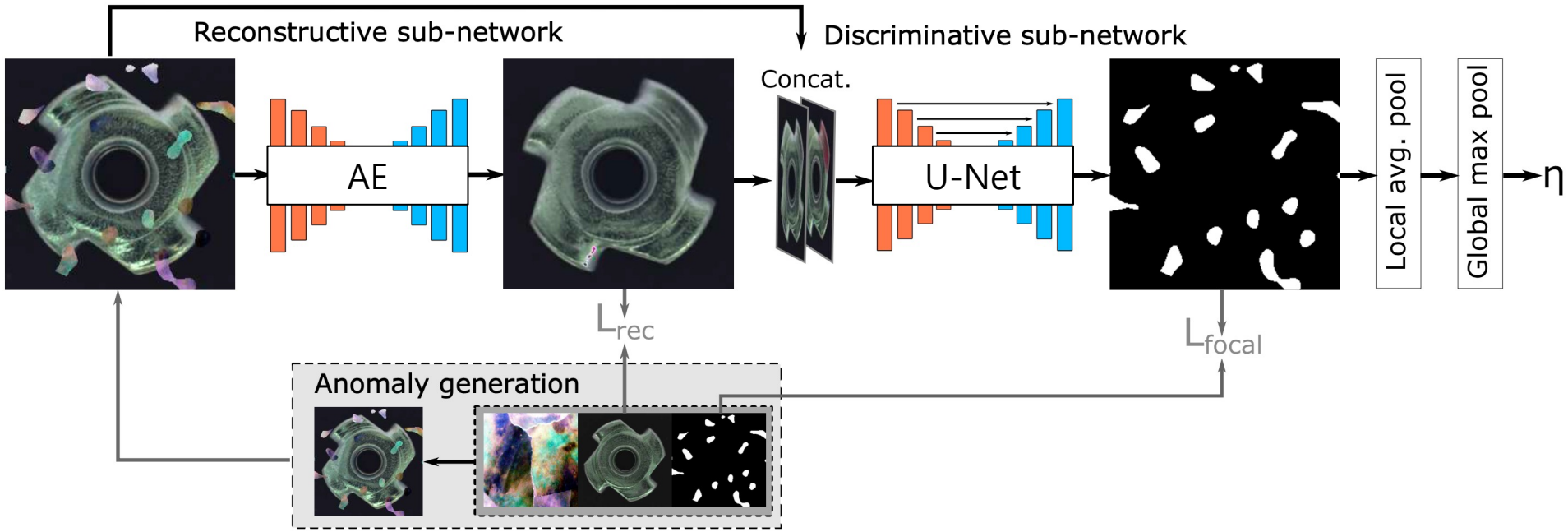


Figure 3. The anomaly detection process of the proposed method. First anomalous regions are implicitly detected and inpainted by the reconstructive sub-network trained using L_{rec}. The output of the reconstructive sub-network and the input image are then concatenated and fed into the discriminative 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.

- **Two sequential networks**
  - **Recon. Net.** : Translate input into anomaly-free form
  - **Dicsr. Net.** : discriminate pixel-wise anomalous from input and reconstruction.

- **One stage end-to-end learning model**

Summary

Proposal

- End-to-End pixel-wise anomaly scoring model
  - Not requires post-processing of output
- Anomalous simulation methods to train neural network
  - Synthesize anomalous on anomaly-free by out-of-distribution patterns
  - Including random opacity β and random augmentation methods

Contributions

- Enables pixel-level anomaly score by end-to-end training
  - Not requires real-world anomalous samples (anomalous simulator)
  - Also, no need for manual annotation
- Shows various experimental results
  - MVTec AD and DAGM dataset
  - Lots of comparison and ablation study

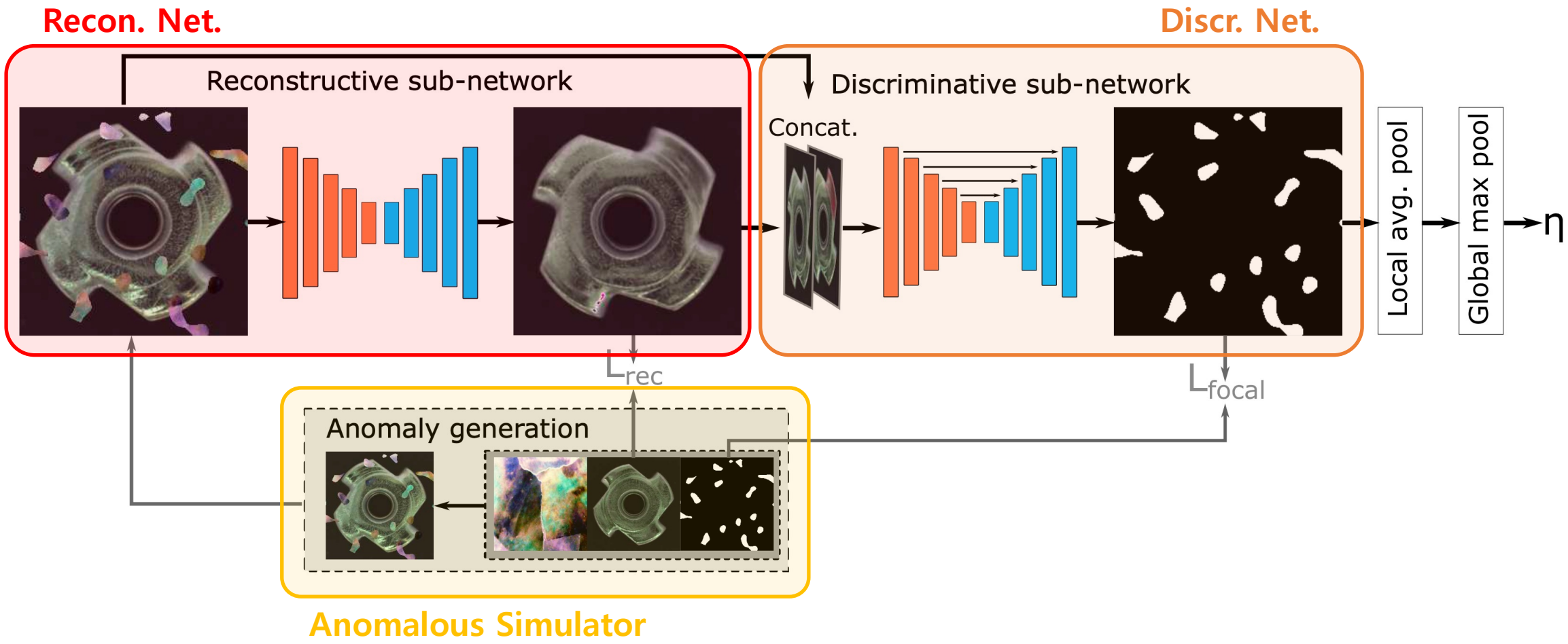
Limitations

- Computational burden due to two sub-networks
  - To apply real-world industry, the scale-up approach should be avoided
- Short explanation (consider the page limit)
  - Why don't use U-Net and MSGMS for reconstruction sub-network?
  - However, the result of using AE with MSGMS is shown
  - Where is the content of the embedding space?

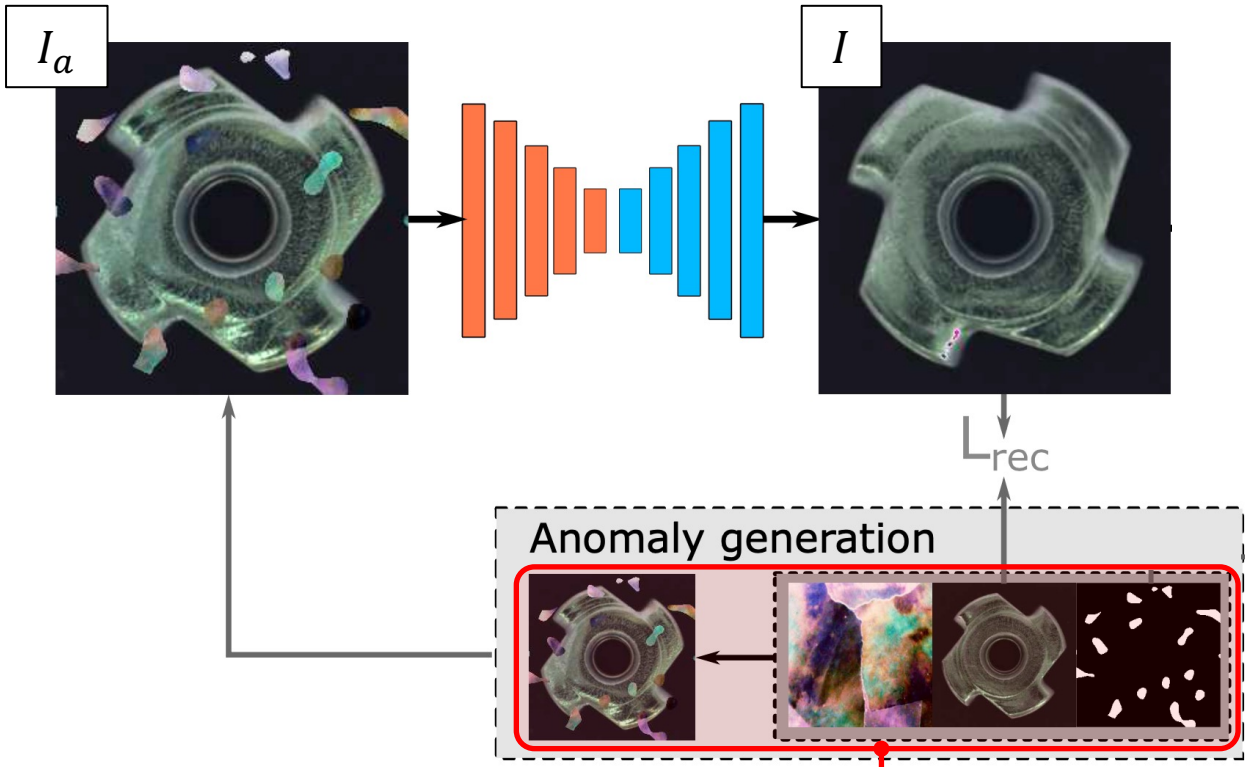
DRÆM

Discriminatively trained Reconstruction Anomaly Embedding Model

DRÆM


Recon. Net.
Discr. Net.
Reconstructive sub-network
Discriminative sub-network
Concat.
Local avg. pool
Global max pool
η
L_{rec}
L_{focal}
Anomaly generation
Anomalous Simulator

Reconstructive Network



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

* β: opacity of anomalous, $\beta = [0.1, 1.0]$

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

* λ: hyper-parameter for loss balancing, $\lambda = 1$

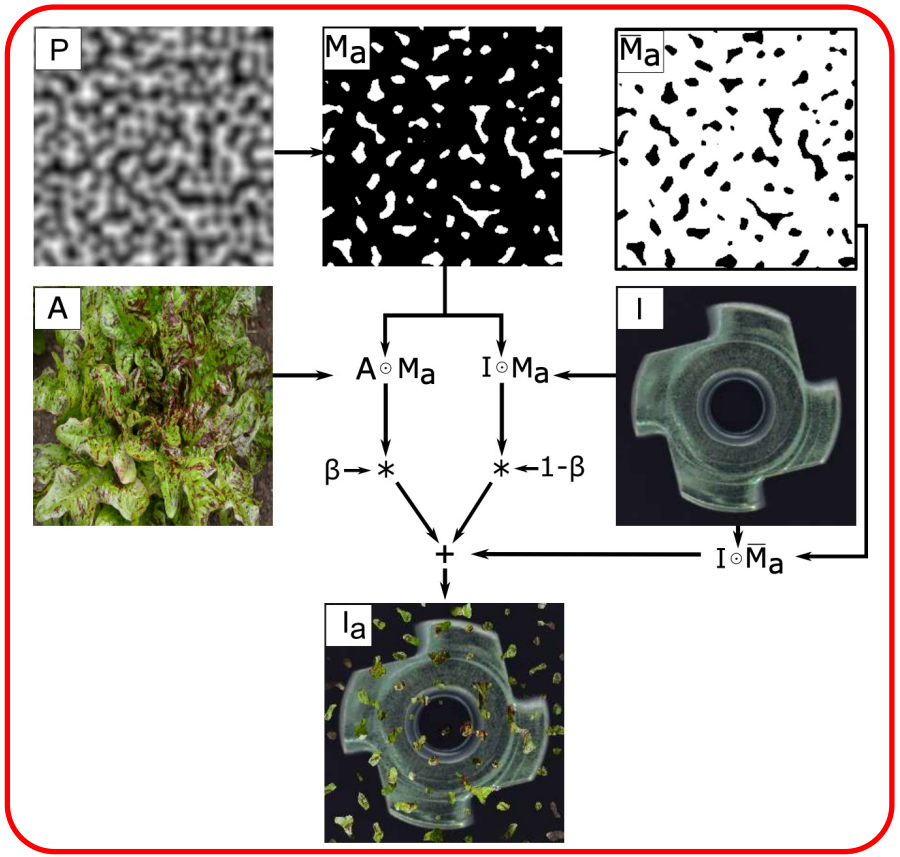
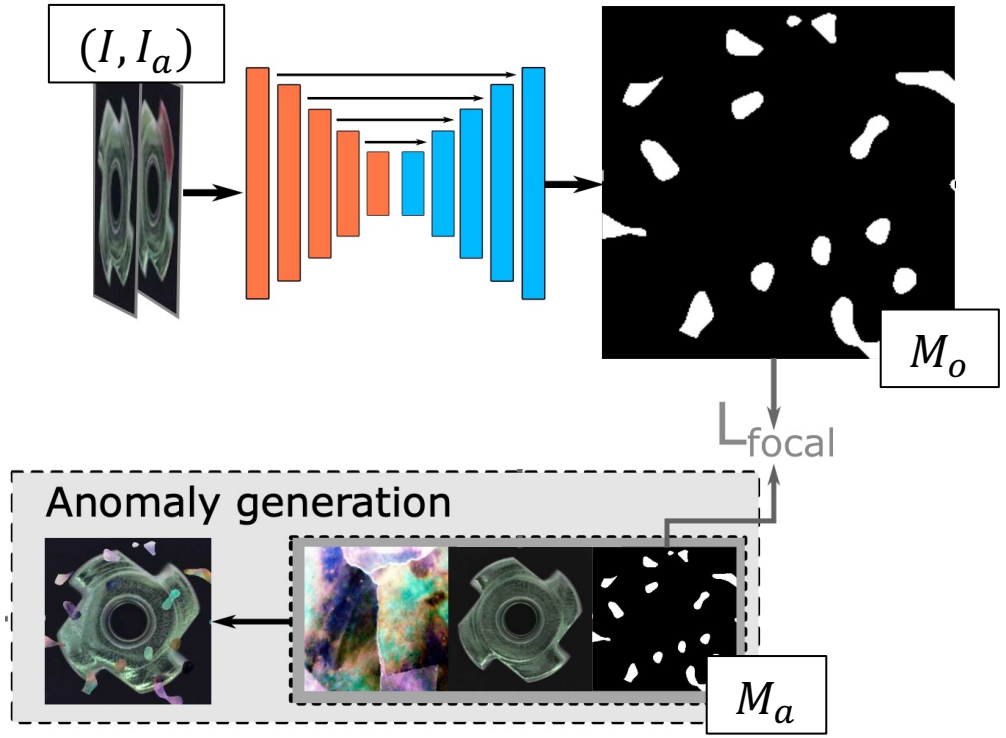


Figure 4. Simulated anomaly generation process. The binary anomaly mask M_a is generated from Perlin noise P. The anomalous regions are sampled from A according to M_a and placed on the anomaly free image I to generate the anomalous image I_a.

Discriminative Network



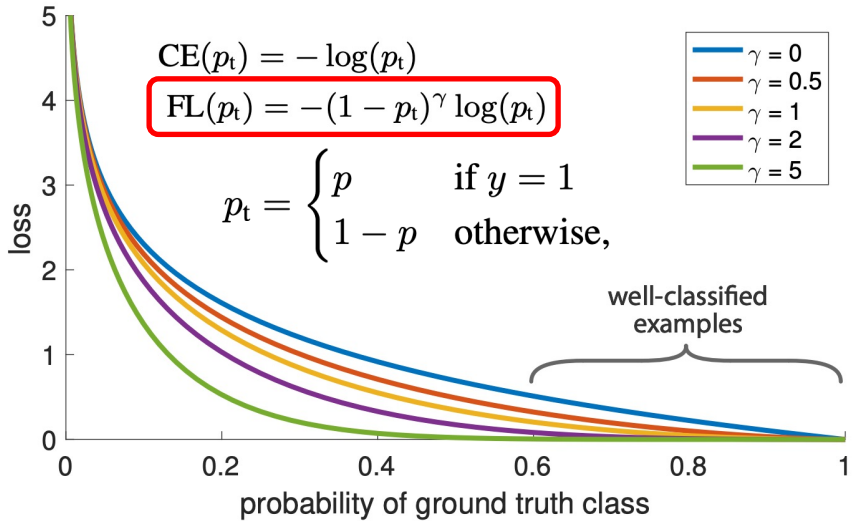


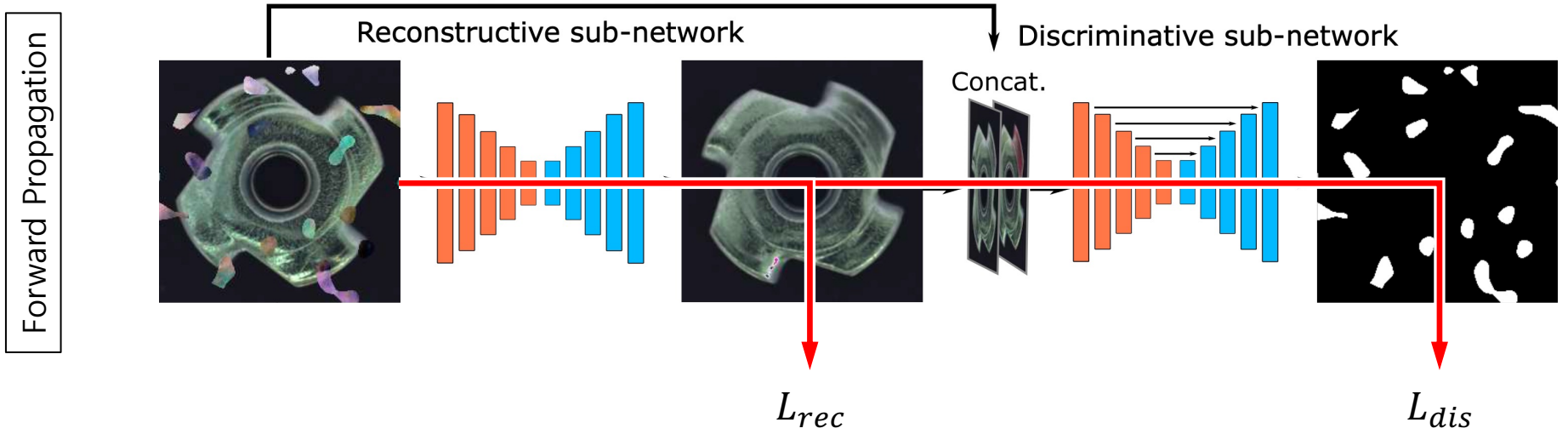
Figure 1. We propose a novel loss we term the *Focal Loss* that adds a factor $(1 - p_t)^{\gamma}$ to the standard cross entropy criterion. Setting $\gamma > 0$ reduces the relative loss for well-classified examples ($p_t > .5$), putting more focus on hard, misclassified examples. As our experiments will demonstrate, the proposed focal loss enables training highly accurate dense object detectors in the presence of vast numbers of easy background examples.

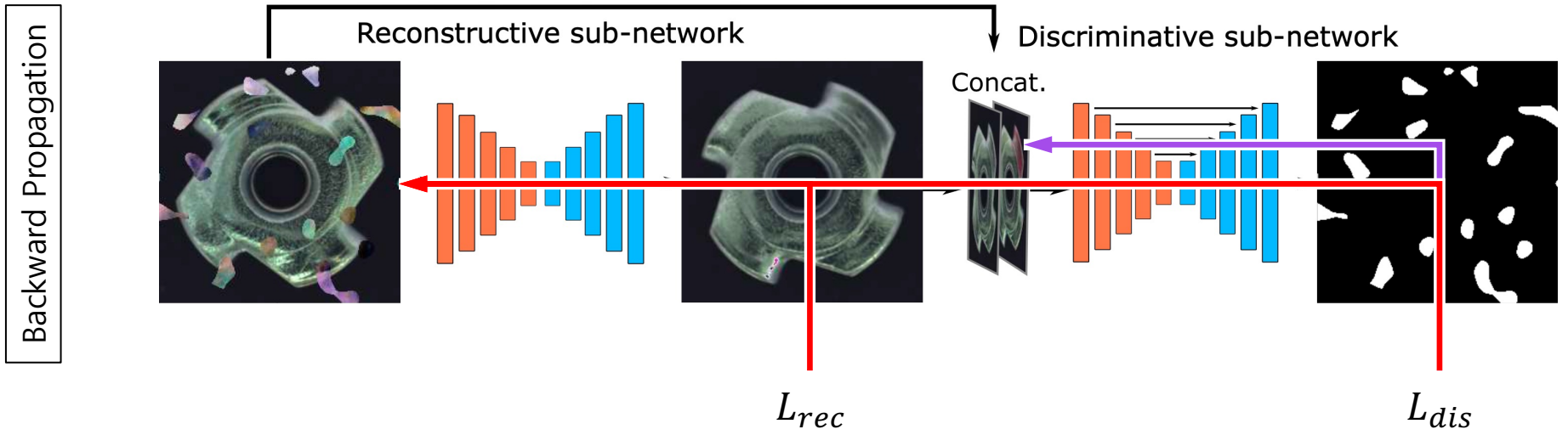
$$L_{dis}(M_a, M_o) = -M_a \log(M_o)$$ Cross-Entropy loss case

$$L_{dis}(M_a, M_o) = -(1 - M_o)^{\gamma} M_a \log(M_o)$$ Focal Loss case

Parameter Update

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





Embedding Space

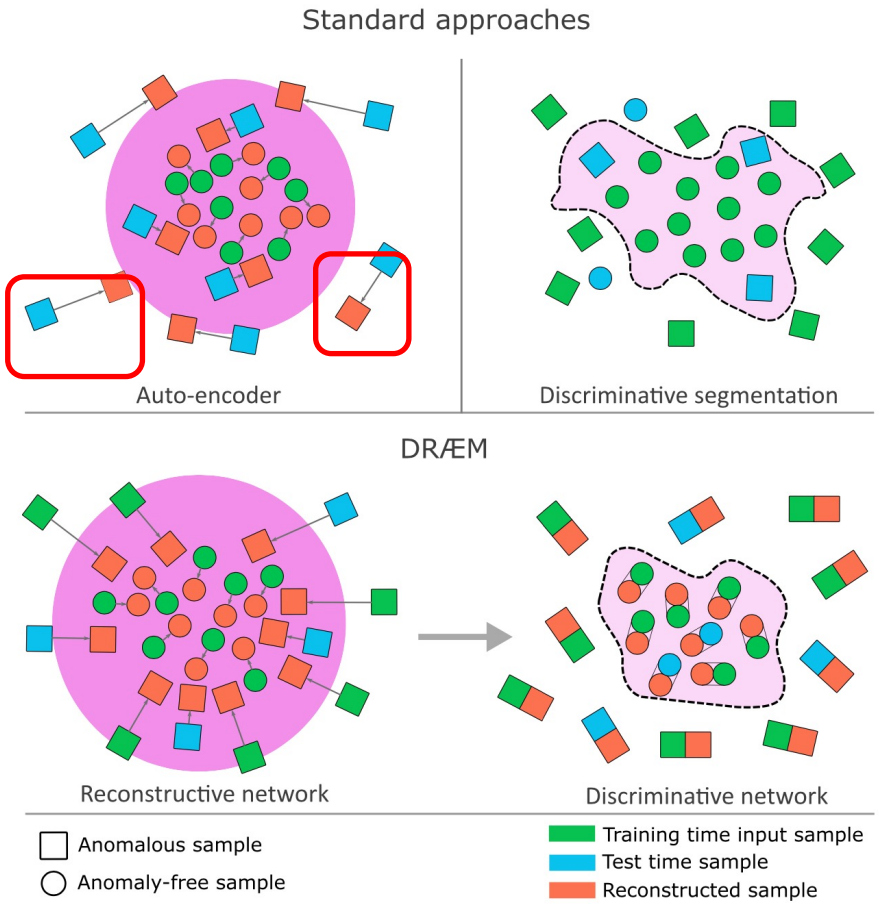
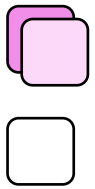


Figure 2. Autoencoders over-generalize to anomalies, while discriminative approaches over-fit to the synthetic anomalies and do not generalize to real data. Our approach jointly discriminatively learns the reconstruction subspace and a hyper-plane over the joint original and reconstructed space using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



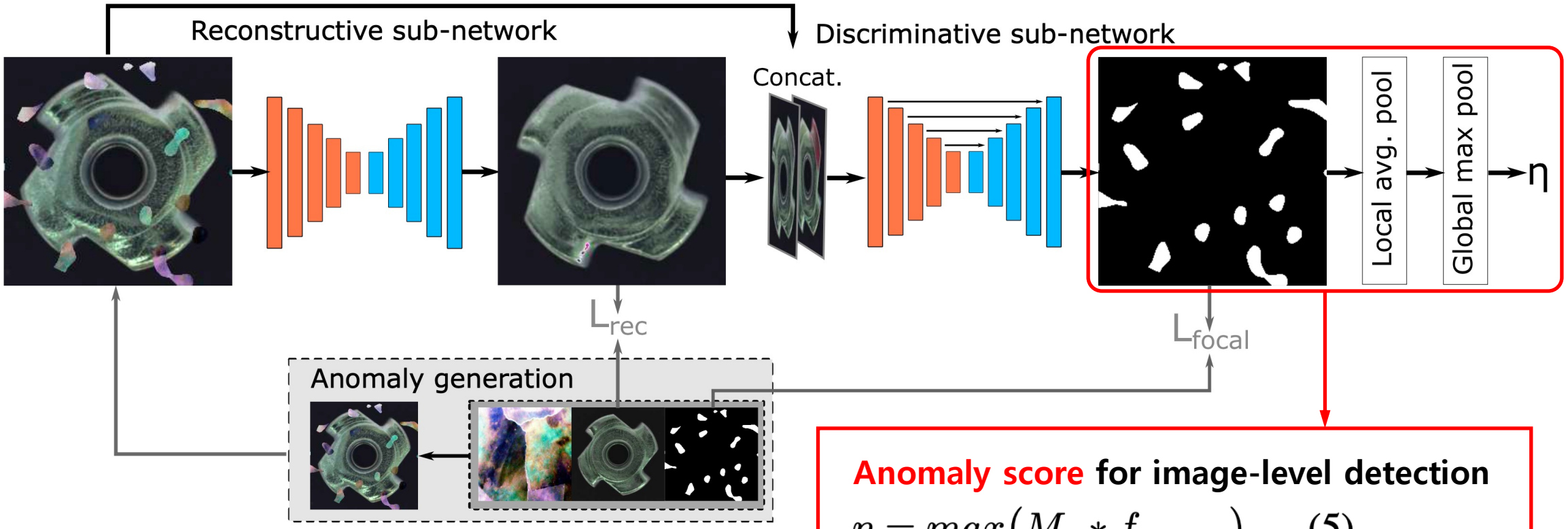
Anomaly-free space

Anomalous space

- Comparison between conventional and DRÆM logically
- **Conventional**
  - AEs (Recon.s) sometimes generalize (well reconstruct) anomalies
  - Discr.s overfit to the training dataset
- **DRÆM**
  - Recon. always translate input into anomaly-free space
  - Discr. shows better generalization (discriminate unseen anomalies)

Experiments

Anomaly Score



Anomaly score for image-level detection

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

$$s_f = 21$$

Qualitative Evaluation

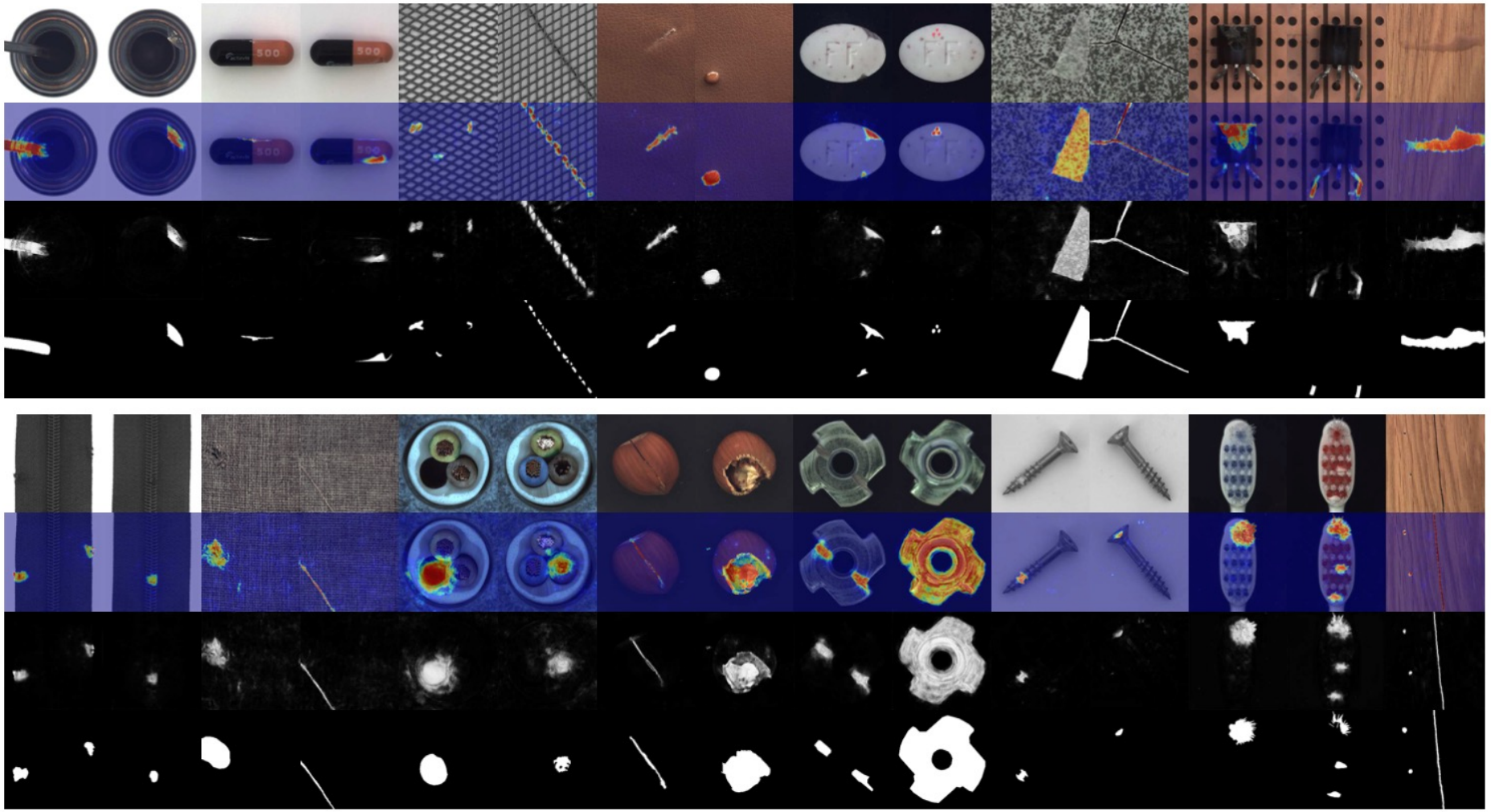


Figure 8. Qualitative examples. The original image, the anomaly map overlay, the anomaly map and the ground truth map are shown.

[4] Bergmann, Paul, et al. "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings." Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2020.
[31] Zavrtanik, Vitjan et al. "Reconstruction by inpainting for visual anomaly detection." *Pattern Recognition* 112 (2021): 107706.
[11] Defard, Thomas, et al. "Padim: a patch distribution modeling framework for anomaly detection and localization." ICPR International Workshops and Challenges, 2021.

Image-Level Detection

Latent Embedding — Patch Distribution Modeling Framework

			US	RIAD		PaDiM	
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**

9 wins

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).

[4] Bergmann, Paul, et al. "Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings." Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition (CVPR). 2020.
[31] Zavrtanik, Vitjan et al. "Reconstruction by inpainting for visual anomaly detection." *Pattern Recognition* 112 (2021): 107706.
[11] Defard, Thomas, et al. "Padim: a patch distribution modeling framework for anomaly detection and localization." ICPR International Workshops and Challenges, 2021.

Pixel-Level Detection

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**

8 / 11 wins

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

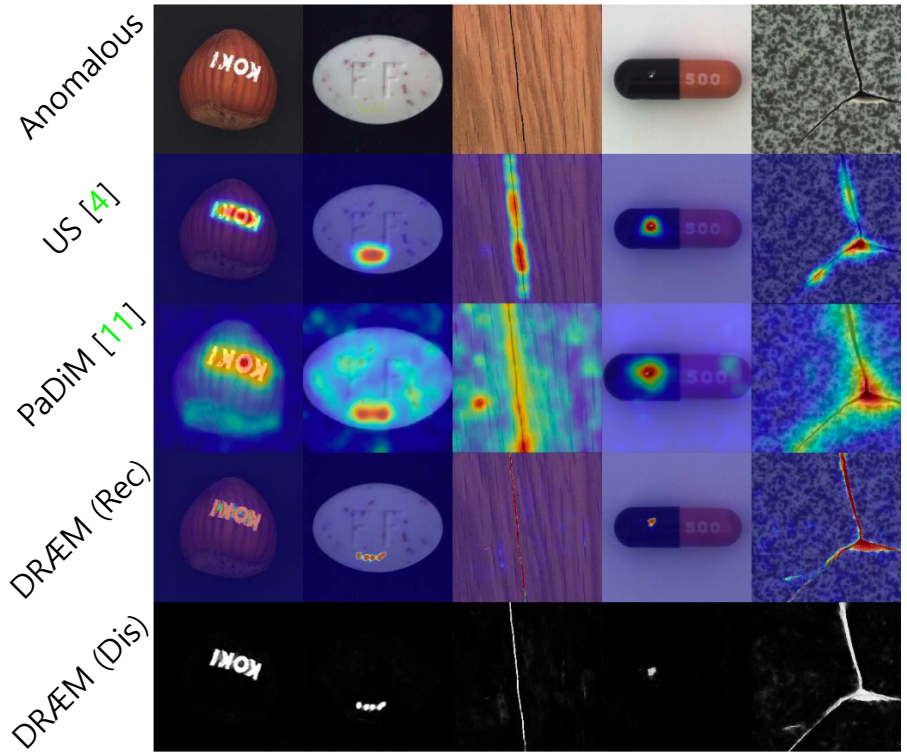


Figure 7. The anomalous images are shown in the first row. The middle three rows show the anomaly maps generated by our implementation of Uninformed Students [4], PaDim [11] and DRÆM, respectively. The last row shows the direct anomaly map output of DRÆM.

Limitations

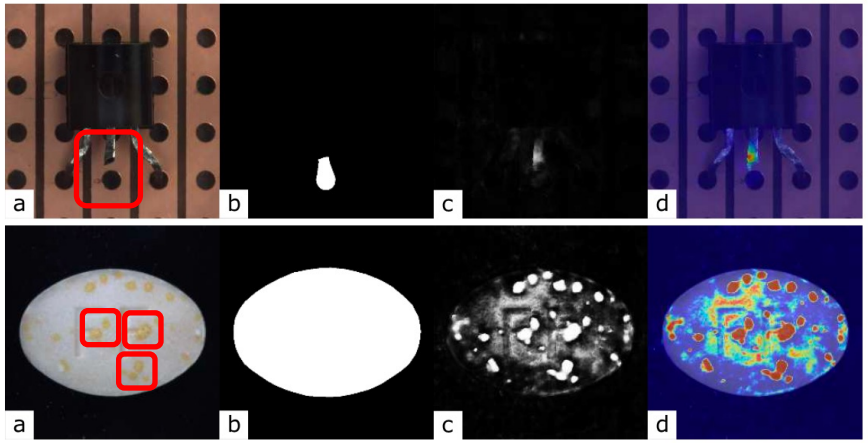


Figure 6. The original image (a) contains anomalies which are difficult to mark in the ground truth mask (b) which causes a discrepancy between the ground truth and the output anomaly map (c,d).

- DRÆM purposes surface anomaly detection
  - Thus, the absence of the part will not be detected well
- Performance was underestimated due to annotation errors

[6] Božič, Jakob et al. "End-to-end training of a two-stage neural network for defect detection." 2020 25th International Conference on Pattern Recognition (ICPR). IEEE, 2021.
[31] Zavrtanik, Vitjan et al. "Reconstruction by inpainting for visual anomaly detection." *Pattern Recognition* 112 (2021): 107706.

Ablation Study

Overall

	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
RIAD [31]	✓							✓	91.7	94.2 / --.-
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.

Ablation Study

Augmentation Perspective

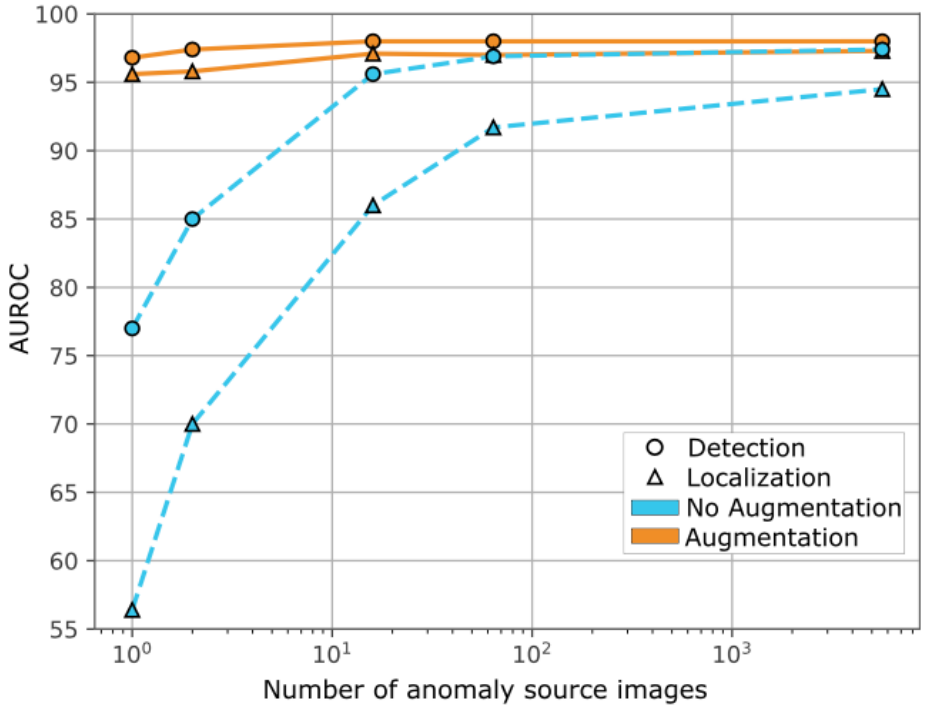


Figure 9. DRÆM achieves a remarkable detection and localization performance already at as low as 10 texture source images in the simulator when augmentation is applied.

In anomalous simulation, the more diverse data makes better performance

Other Dataset

DAGM dataset

	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.

- DAGM dataset provides GT for training

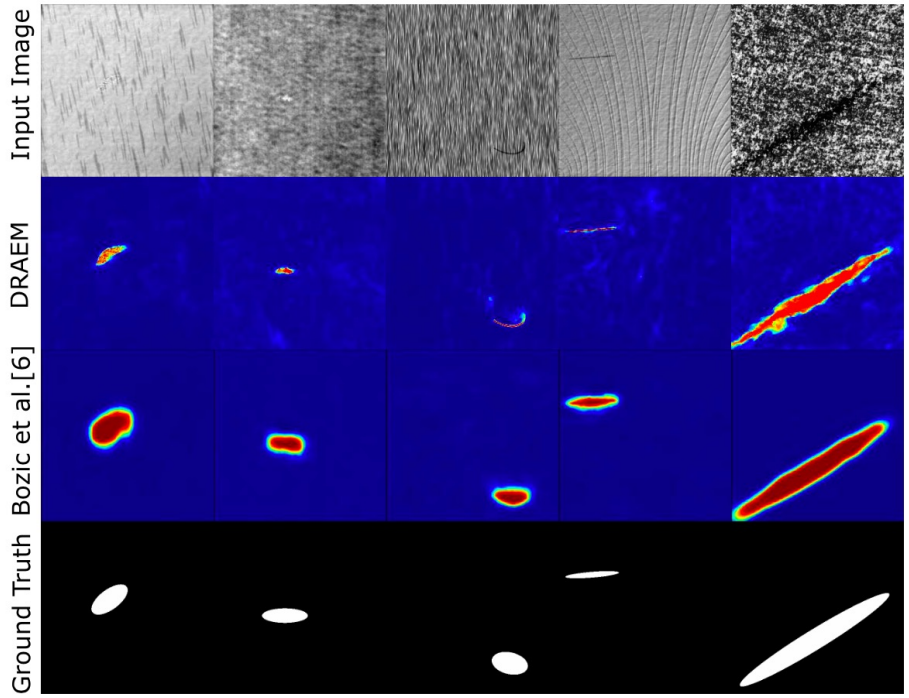


Figure 11. Supervised methods replicate the approximate ground truth training annotations, leading to a low localization accuracy. DRÆM does not use the ground truth, yet produces far better localization.

Review Comment

Few confusing or odd points

- Still symbol notation mismatch
  - L_{seg} and L_{focal}: loss function of discriminative sub-network
  - M and M_o: output of the discriminative sub-network

- A detailed analysis of the technique is short
  - Why use AE than U-Net for reconstructive sub-network?
  - Why don't use MSGMS for reconstructive sub-network?

But some acceptable points are exist!

- Anomalous simulation method is acceptable
  - It can train reconstruction sub-network to translate from anomalous to anomaly-free
  - Also, authors show the effect of anomalous simulation (w/ augmentation, large-scale dataset for synthesizing)

- Discriminative sub-network enables anomalous localization in end-to-end manner
  - It is designed to eliminate post-processing
  - But it still requires post-processing again

Appendix A.

DSR

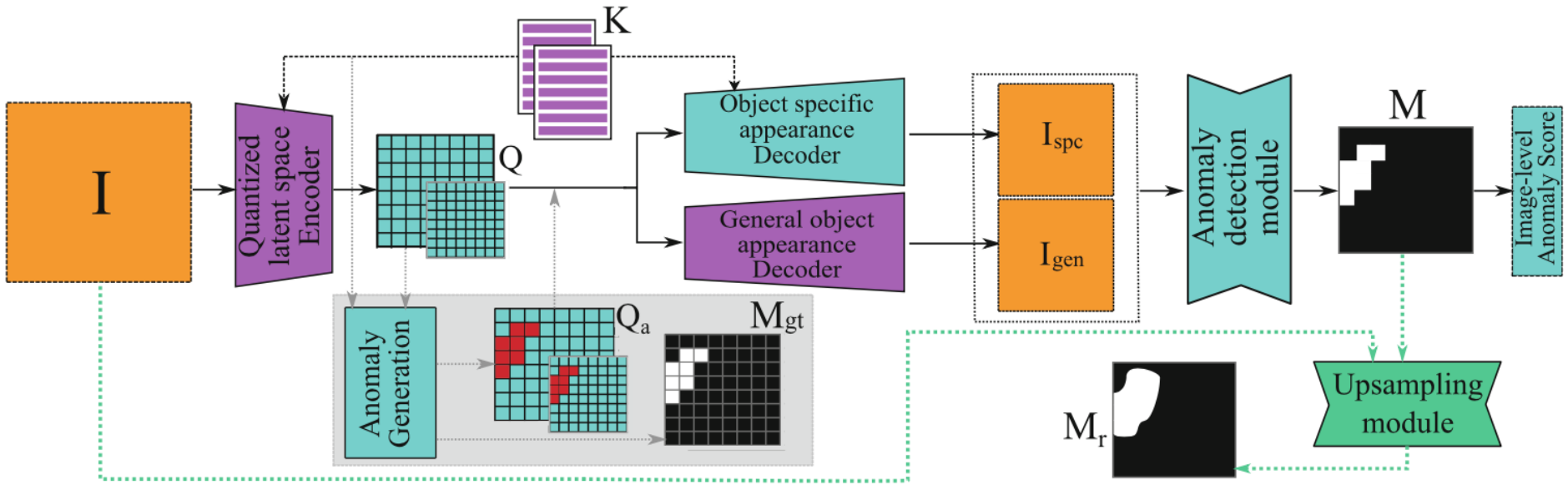


Fig. 2. The DSR architecture. During training, the non-anomalous image quantized feature maps (Q_{hi}, Q_{lo}) are replaced by the anomaly augmented feature maps $(Q_{a,hi}, Q_{a,lo})$ generated by the latent space sampling procedure (shaded block). The pathway marked with green arrows are used when training the Upsampling module with simulated smudges and at inference.

DSR

Table 3. Results of anomaly detection on MVTec dataset (AUROC) with the average score over all classes (*avg*) in the last column.

	Method	bottle	capsule	grid	leather	pill	tile	trans.	zipper	cable	carpet	hazelnut	m. nut	screw	toothbrush	wood	average
US	[4]	99.0	86.1	81.0	88.2	87.9	99.1	81.8	91.9	86.2	91.6	93.1	82.0	54.9	95.3	97.7	87.7
RIAD	[22]	99.9	88.4	99.6	**100**	83.8	98.7	90.9	98.1	81.9	84.2	83.3	88.5	84.5	**100**	93.0	91.7
	[17]	**100**	92.3	92.9	**100**	83.3	97.4	95.9	97.9	**94.0**	95.5	98.7	93.1	81.2	95.8	97.6	94.4
PaDiM	[7]	99.8	91.5	95.7	**100**	94.4	97.4	**97.8**	90.9	92.2	99.9	93.3	99.2	84.4	97.2	98.8	95.5
CutPaste	[11]	98.2	98.2	**100**	**100**	94.9	94.6	96.1	99.9	81.2	93.9	98.3	**99.9**	88.7	99.4	**99.1**	96.1
DRÆM	[21]	99.2	**98.5**	99.9	**100**	**98.9**	99.6	93.1	**100**	91.8	97.0	**100**	98.7	93.9	**100**	**99.1**	98.0
	DSR	**100**	98.1	**100**	**100**	97.5	**100**	**97.8**	**100**	93.8	**100**	95.6	98.5	**96.2**	99.7	96.3	**98.2**

Results of KolektorSDD2

Table 1. Anomaly detection (AP_{det}) and localization (AP_{loc}) on the KSDD2 dataset.

Method	US [4]	MAD [17]	DRÆM [21]	PaDim [7]	DSR	AMAD	$AMAD_P$
AP_{det}	65.3	79.3	77.8	55.6	**87.2**	82.9	84.0
AP_{loc}	-	-	42.4	45.3	**61.4**	-	-

Appendix B.

OCR-GAN

Omni-frequency Channel-selection Reconstruction

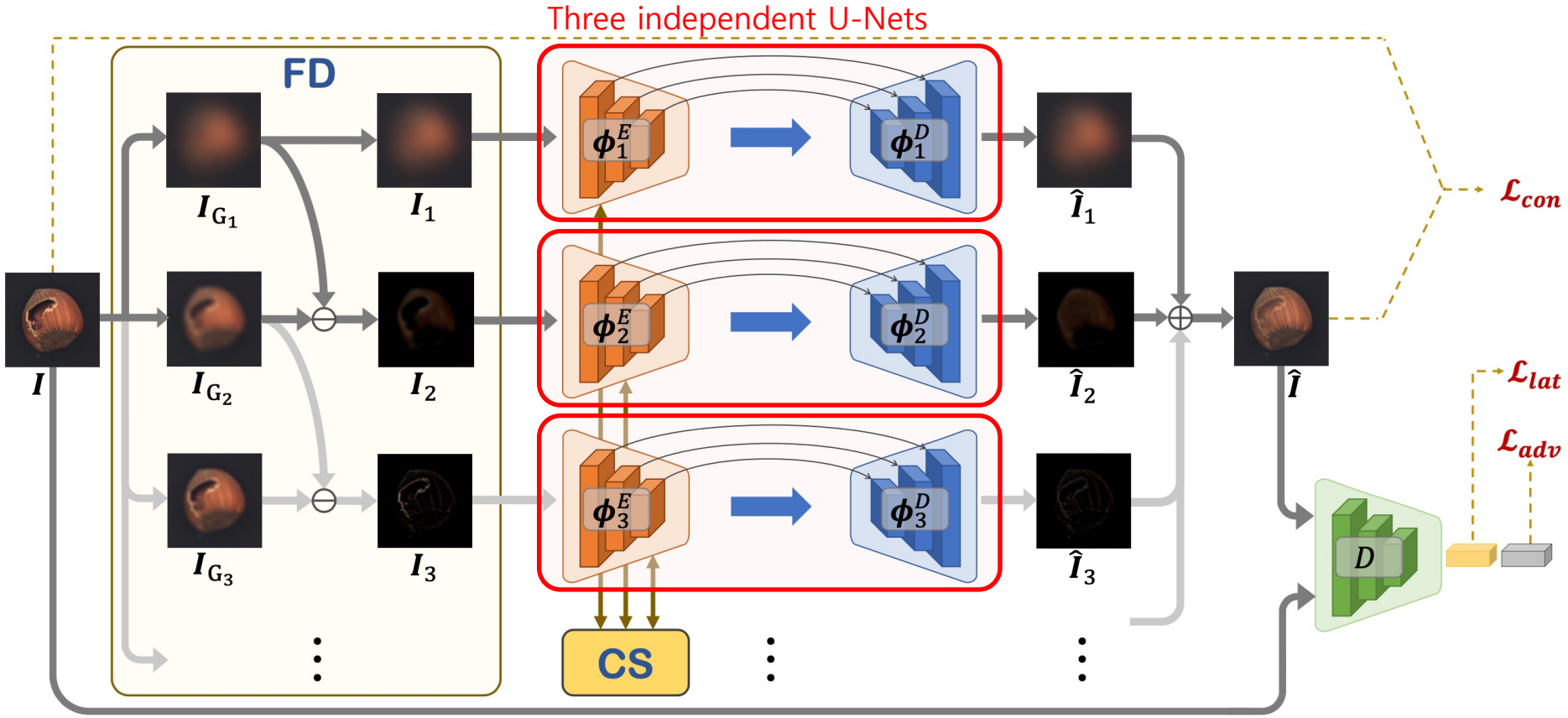


Fig. 4. **Overview of proposed OCR-GAN**. Input image $\boldsymbol{I}$ goes through Frequency Decoupling (FD) module to obtain omni-frequency images $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \dots\}$ from pre-processed Gaussian images $\{\boldsymbol{I}_{G_1}, \boldsymbol{I}_{G_2}, \dots\}$. Then $\{\boldsymbol{I}_1, \boldsymbol{I}_2, \dots\}$ are fed into multiple generators $\{\phi_1, \phi_2, \dots\}$ to reconstruct corresponding images $\{\hat{\boldsymbol{I}}_1, \hat{\boldsymbol{I}}_2, \dots\}$, which are added to obtain the final output $\hat{\boldsymbol{I}}$. The proposed Channel Selection (CS) module performs omni-frequency interaction among different encoders, *i.e.*, $\{\phi_1^E, \phi_2^E, \dots\}$.

Different frequency branches in the framework are independent by default.

OCR-GAN

extra training data. ***c)*** *Reconstruction-based methods* [15]–[18] contain a generator to reconstruct the input image, and the anomaly score is the more interpretable reconstruction error. *These methods do not need pre-trained models and extra training data*. However, current reconstruction-based methods without extra training data are much less expressive than other methods for the generator's poor reconstruction ability. In summary, current unsupervised anomaly detection approaches are still suffering from two main challenges: ***(1))*** Some works achieve high AUC score but require abnormal samples or extra training data that are hard to obtain and costly for practical use. ***(2))*** Current reconstruction-based methods are more practical and do not need pre-trained models and extra training data but suffer from low performance. This paper focuses on studying the reconstruction-based method as it requires no extra training data and only normal samples that is more practical.

Liang, Yufei, et al. "Omni-frequency channel-selection representations for unsupervised anomaly detection." arXiv preprint arXiv:2203.00259 (2022).

OCR-GAN

TABLE I

AUC RESULTS WITH SOTAS ON MVTEC AD DATASET. THREE TO TEN COLUMNS ARE RECONSTRUCTION-BASED METHODS WHILE THE FOLLOWING FOUR COLUMNS ARE DENSITY-BASED AND CLASSIFICATION-BASED METHODS. **BOLD** AND UNDERLINE REPRESENT OPTIMAL AND SUBOPTIMAL RESULTS. ' MEANS THE YEAR OF PUBLICATION. † MEANS USING THE PRE-TRAINED MODEL WITH EXTRA DATASET. ‡ MEANS OUR TRAINING WITH FORGERY ABNORMAL SAMPLES IN SECTION IV-C.

	Items	AGAN [59] 17'	AE_1 [48] 18'	AE_2 [48] 18'	SkipG [21] 19'	GradC [55] 20'	P-AE [61] 20'	DGAD [17] 21'	Draem [19] 21'	Diff [39] 21'	CutPaste [14] 21'	CutPaste† [14] 21'	InTra [62] 21'	Ours	Ours‡
texture	Carpet	49.0	67.0	50.0	70.9	89.3	65.7	52.0	97.0	92.9	93.1	**100.0**	98.8	98.9±0.5	99.4±0.3
	Grid	51.0	69.0	78.0	47.7	71.6	75.4	67.0	99.9	84.0	99.9	99.1	**100.0**	99.6±0.2	99.6±0.2
	Leather	52.0	46.0	44.0	60.9	69.3	72.9	94.0	**100.0**	97.1	**100.0**	**100.0**	**100.0**	97.1±0.6	97.1±0.8
	Tile	51.0	52.0	77.0	29.9	63.4	65.5	83.0	99.6	99.4	93.4	**99.8**	98.2	92.2±0.8	95.5±1.5
	Wood	68.0	83.0	74.0	19.9	76.7	89.5	72.0	99.1	**99.8**	98.6	**99.8**	98.0	95.8±1.6	95.7±1.1
	Average	54.2	63.4	64.6	45.86	74.1	73.8	73.6	99.1	94.6	95.7	**99.7**	99.0	96.6±0.3	97.5±0.3
object	Bottle	69.0	88.0	80.0	85.2	52.0	94.2	97.0	99.2	99.0	98.3	**100.0**	**100.0**	99.6±0.2	99.6±0.1
	Cable	53.0	61.0	56.0	54.4	58.7	87.9	90.0	91.8	95.9	80.6	96.2	84.2	**99.2**±0.5	99.1±0.6
	Capsule	58.0	61.0	62.0	54.3	55.6	66.9	60.0	**98.5**	86.9	96.2	95.4	86.5	95.4±0.4	96.2±0.6
	Hazelnut	50.0	54.0	88.0	24.5	91.4	91.2	80.0	**100.0**	99.3	97.3	99.9	95.7	88.2±2.0	98.5±1.3
	Metal Nut	50.0	54.0	73.0	81.4	56.0	66.3	95.0	98.7	96.1	99.3	98.6	96.9	98.7±0.2	**99.5**±0.3
	Pill	62.0	60.0	62.0	67.1	92.4	71.6	76.0	**98.9**	88.8	64.7	93.3	90.2	98.5±0.4	98.3±0.2
	Screw	35.0	51.0	69.0	87.9	78.2	57.8	67.0	93.9	96.3	86.3	86.6	95.7	**100.0**±0.0	**100.0**±0.0
	Toothbrush	57.0	74.0	98.0	58.6	98.0	97.8	93.0	**100.0**	98.6	98.3	90.7	**99.7**	98.2±0.9	98.7±0.7
	Transistor	67.0	52.0	71.0	84.5	72.8	86.0	88.0	93.1	91.1	95.5	97.5	95.8	94.9±0.3	**98.3**±1.5
	Zipper	59.0	80.0	80.0	76.1	56.6	75.7	82.0	**100**	95.1	99.4	99.9	99.4	97.6±0.4	99.0±0.2
	Average	56.0	63.5	73.9	67.4	71.2	79.5	82.8	97.4	94.7	94.3	95.8	94.4	97.0±0.2	**98.7**±0.3
	All	55.0	63.0	71.0	60.2	72.1	77.6	80.0	98.0	94.7	95.2	97.1	95.9	96.9±0.2	**98.3**±0.2

OCR-GAN

TABLE VI
ABLATION STUDY FOR FREQUENCY BRANCHES.

Category	high frequency	low frequency	two branches	OCR-GAN
texture	85.2	69.8	73.6	96.6
object	81.3	75.1	75.4	97.0
all	82.6	73.3	74.8	96.9