Paper Review

Reconstruction by Inpainting for Visual Anomaly Detection (RIAD)

YeongHyeon Park

Department of Electrical and Computer Engineering

SungKyunKwan University

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.


UNIVERSITY
OF LJUBLJANA
among top 3 % in the world

Warm UP

Wang, Zhou, et al. "Image quality assessment: from error visibility to structural similarity." IEEE Transactions on Image Processing (TIP), vol.13, 2004.
41101 citations

Structural Similarity

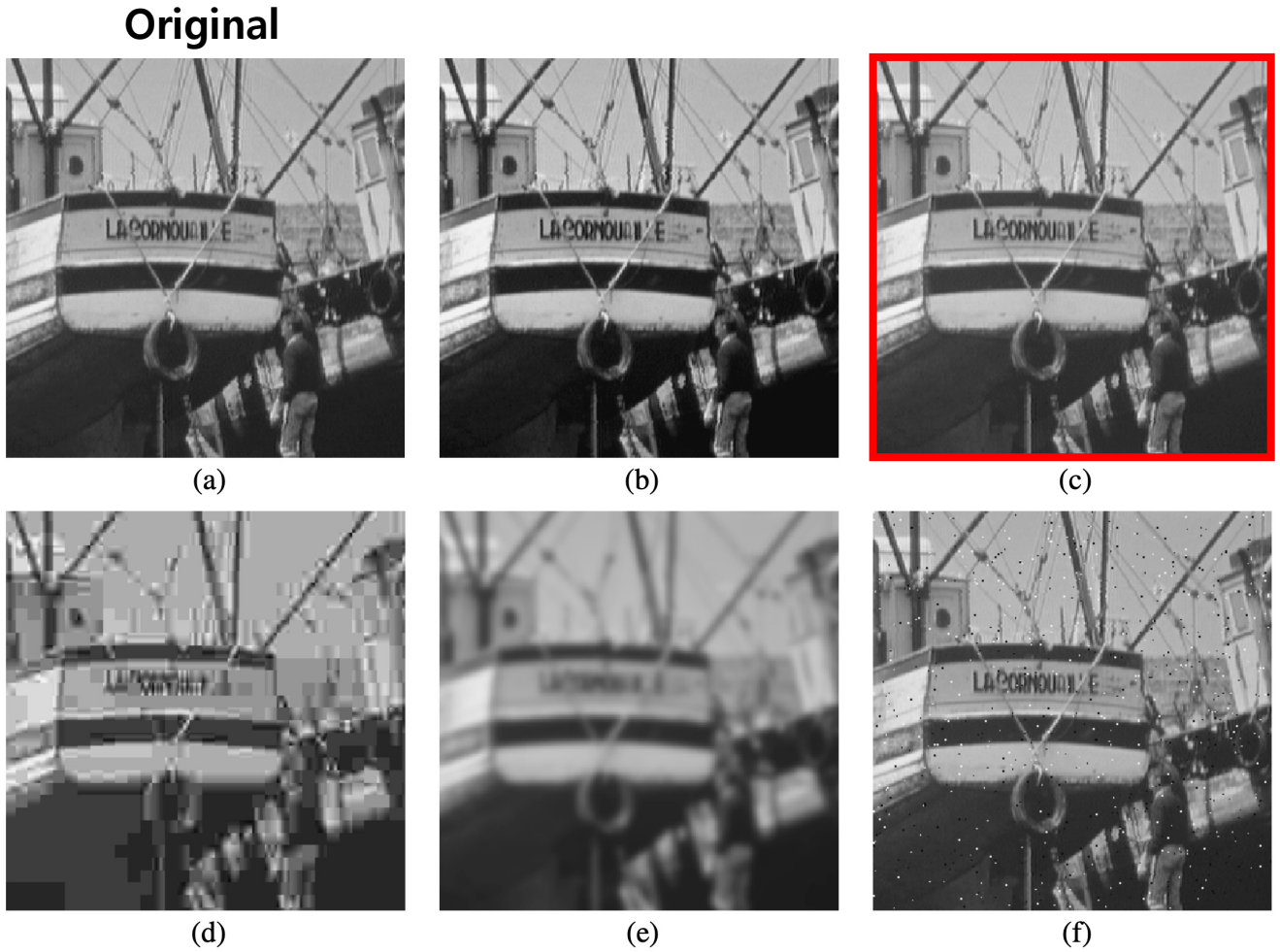


Fig. 2. Comparison of "Boat" images with different types of distortions, all with $\mathrm{MSE} = 210$. (a) Original image (8 bits/pixel; cropped from 512 × 512 to 256 × 256 for visibility). (b) Contrast-stretched image, $\mathrm{MSSIM} = 0.9168$. (c) Mean-shifted image, $\mathrm{MSSIM} = 0.9900$. (d) JPEG compressed image, $\mathrm{MSSIM} = 0.6949$. (e) Blurred image, $\mathrm{MSSIM} = 0.7052$. (f) Salt-pepper impulsive noise contaminated image, $\mathrm{MSSIM} = 0.7748$.

Structural Similarity

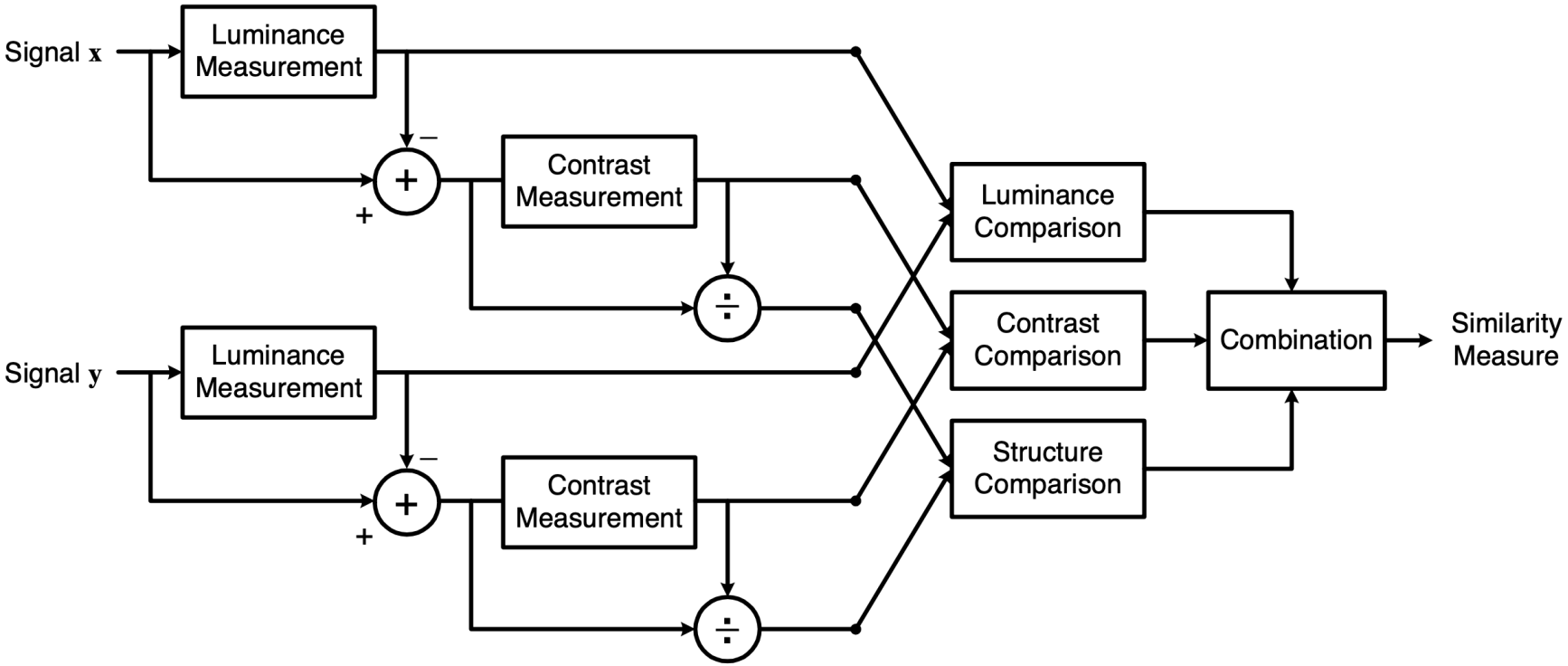


Fig. 3. Diagram of the structural similarity (SSIM) measurement system.

Structural Similarity

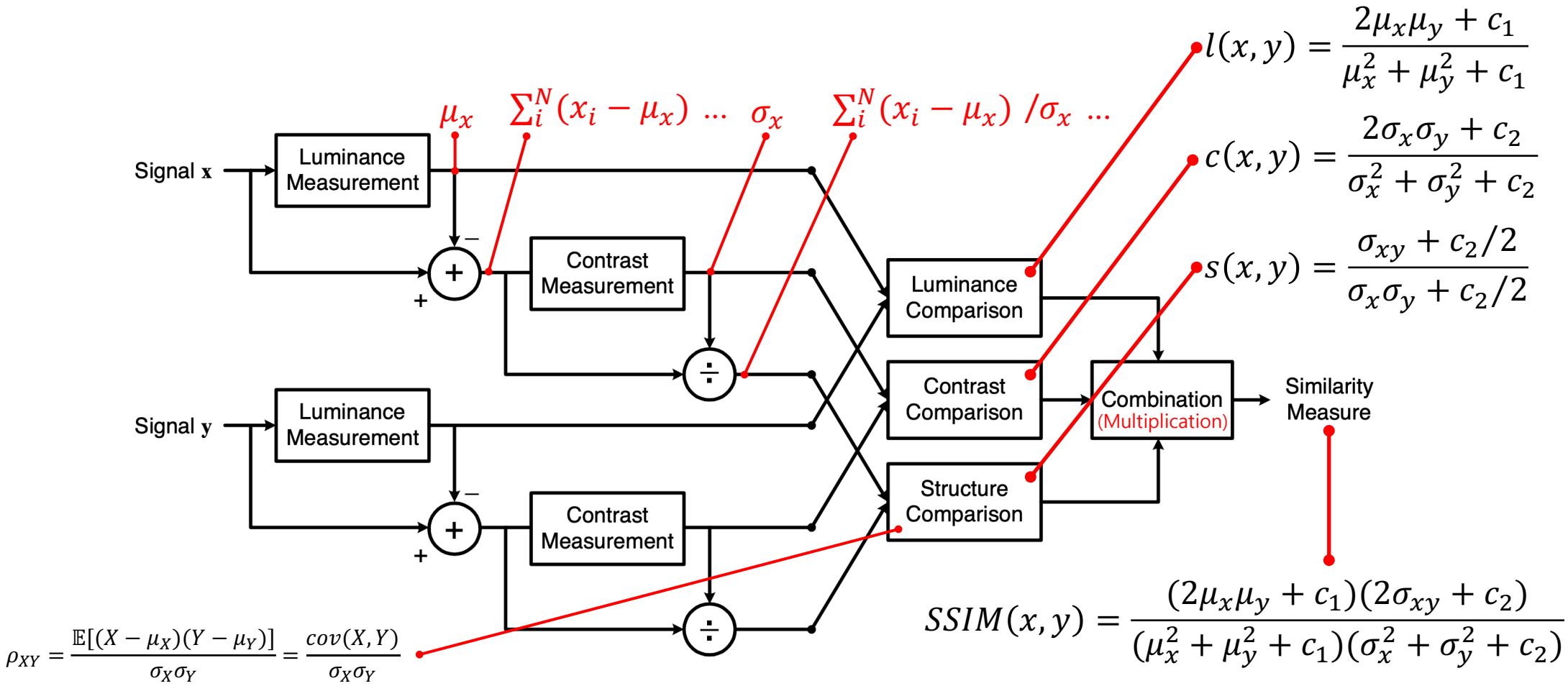


Fig. 3. Diagram of the structural similarity (SSIM) measurement system.

RIAD

Reconstruction by **I**npainting for Visual **A**nomaly **D**etection

Concept

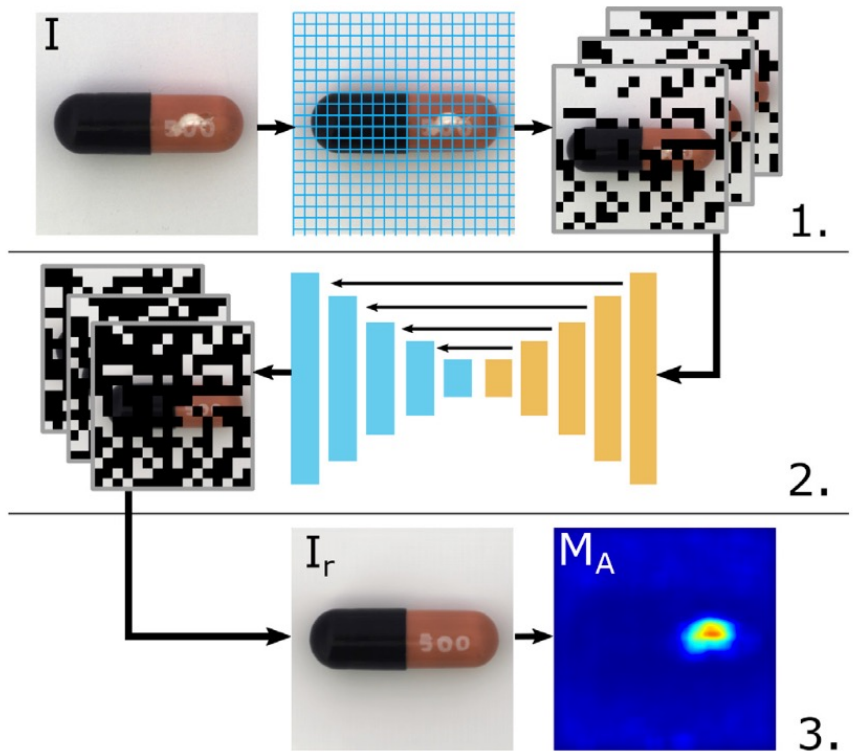


Fig. 1. Our anomaly detection method is based on reconstruction-by-inpainting. The image **I** is split into a grid of rectangular regions of the size of $k \times k$ pixels. The set of rectangular regions is randomly split into n disjoint subsets. For each subset, the regions belonging to that subset are removed from the original image (they are set to 0), resulting in n input images (1). The input images are reconstructed by an inpainting network generating n output images, each reconstructing the regions removed in their corresponding input image (2). The individual reconstructed regions from the n partial reconstructions are re-assembled into a single reconstructed image $\mathbf{I}_r$. The reconstruction quality is then evaluated to generate an anomaly map $M_A(3)$.

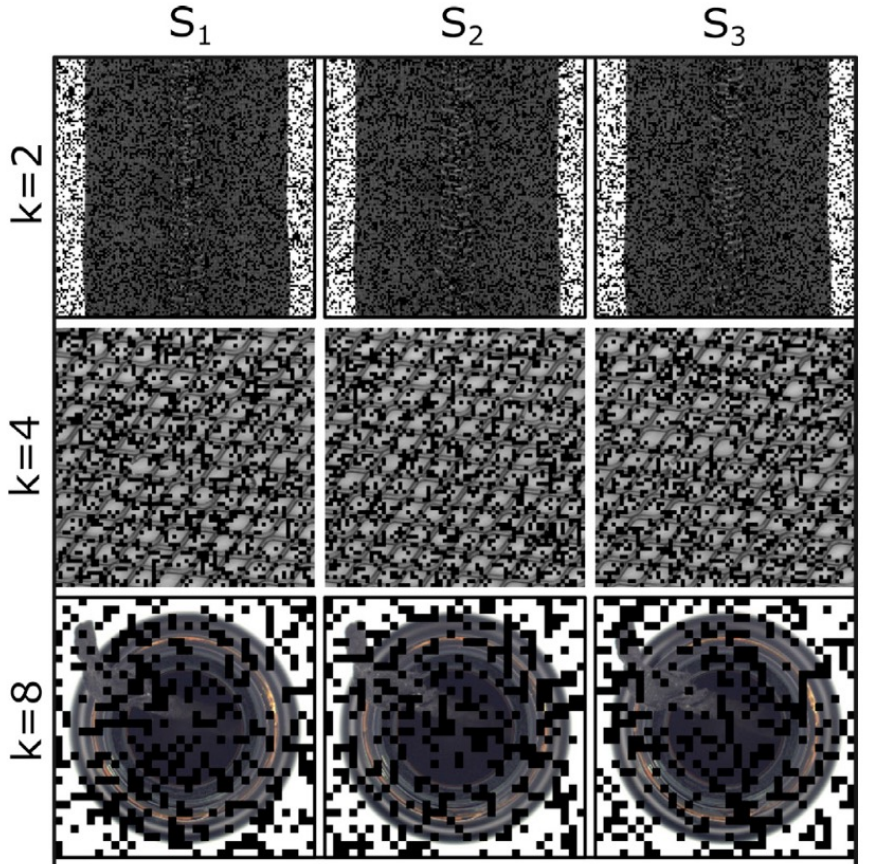


Fig. 2. Examples of an input image masked by $n = 3$ disjoint sets of inpainting regions with masked region sizes $k \in \{2, 4, 8, 16\}$. Note that in every row of images, which show masked images using different masked region sizes, each pixel in the image is masked exactly once.

- Random Cell Masking (left)
- Multi-scale Masking (right)

Summary of RIAD

Proposal

- Patch inpainting-based anomaly detection method.
  - Utilizing a characteristic of the low likelihood between anomaly and neighbor (normal) patches.
- Multi-scale cell masking method for inpainting-based anomaly detection.
- Gradient Magnitude Similarity (GMS) to improve the Structural similarity (SSIM) measure
  - Gradient Magnitude Similarity = Edge similarity

Contributions

- RIAD detects various-sized defective via multi-scale cell masking.
  - It reduces the effort of finding a cell size to cover various-sized defectives properly.
- RIAD does not suffer from the over-detection and miss-detection via multi-resolution inference.
  - Over-detection: judge normal as abnormal (eased in low-resolution comparison)
  - Miss-detection judge abnormal as normal (eased in low-resolution comparison)
- GMS enhances the robustness of SSIM measure.

Limitations

- RIAD needs a lot of time for inference by double loop operation.
  - Loop 1 for multi-scale cell masking ($k \in K$)
  - Loop 2 for multi-resolution inference ($l \in L$)
  - Parallelization will be possible but needs a powerful machine.
- Since RIAD performs via random masking, sometimes anomalies may not be detected well.
  - Lottery anomaly detection model

Summary of RIAD

Overall good!

- Logical statements for the proposed method.
- Various comparative experiments for verification.

- However, some symbols are not matched or confused.

Anomaly Detection with RIAD

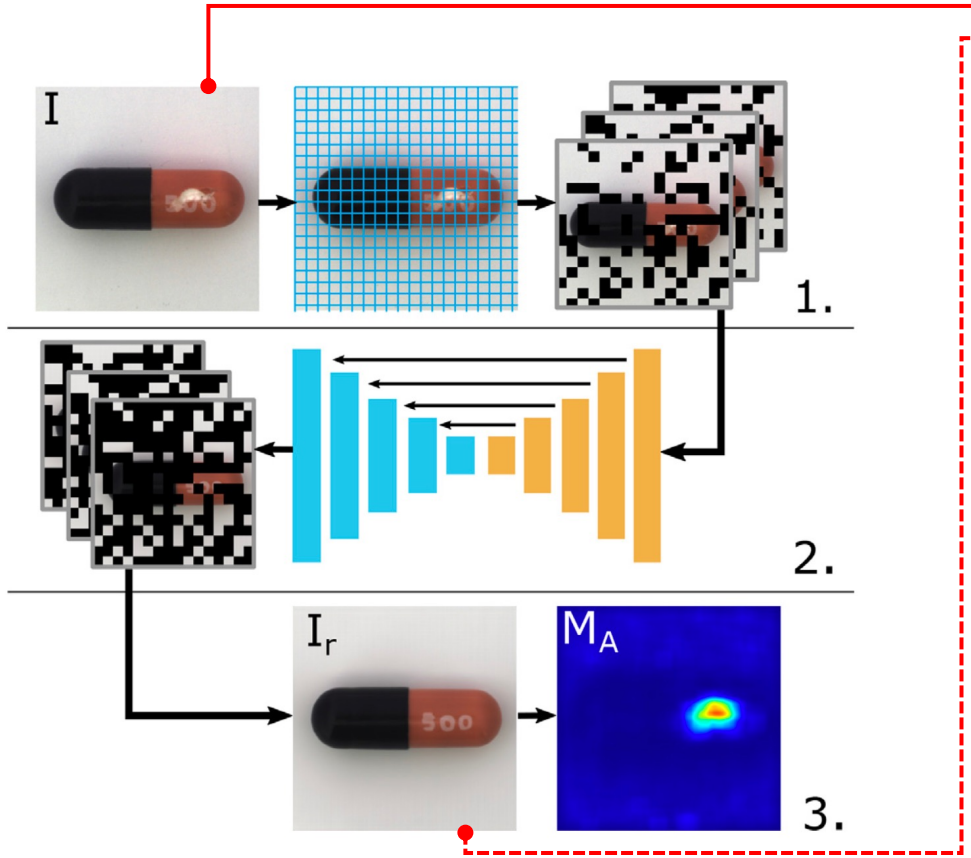


Fig. 1. Our anomaly detection method is based on reconstruction-by-inpainting. The image **I** is split into a grid of rectangular regions of the size of $k \times k$ pixels. The set of rectangular regions is randomly split into n disjoint subsets. For each subset, the regions belonging to that subset are removed from the original image (they are set to 0), resulting in n input images (1). The input images are reconstructed by an inpainting network generating n output images, each reconstructing the regions removed in their corresponding input image (2). The individual reconstructed regions from the n partial reconstructions are re-assembled into a single reconstructed image $\mathbf{I}_r$. The reconstruction quality is then evaluated to generate an anomaly map M_A(3).

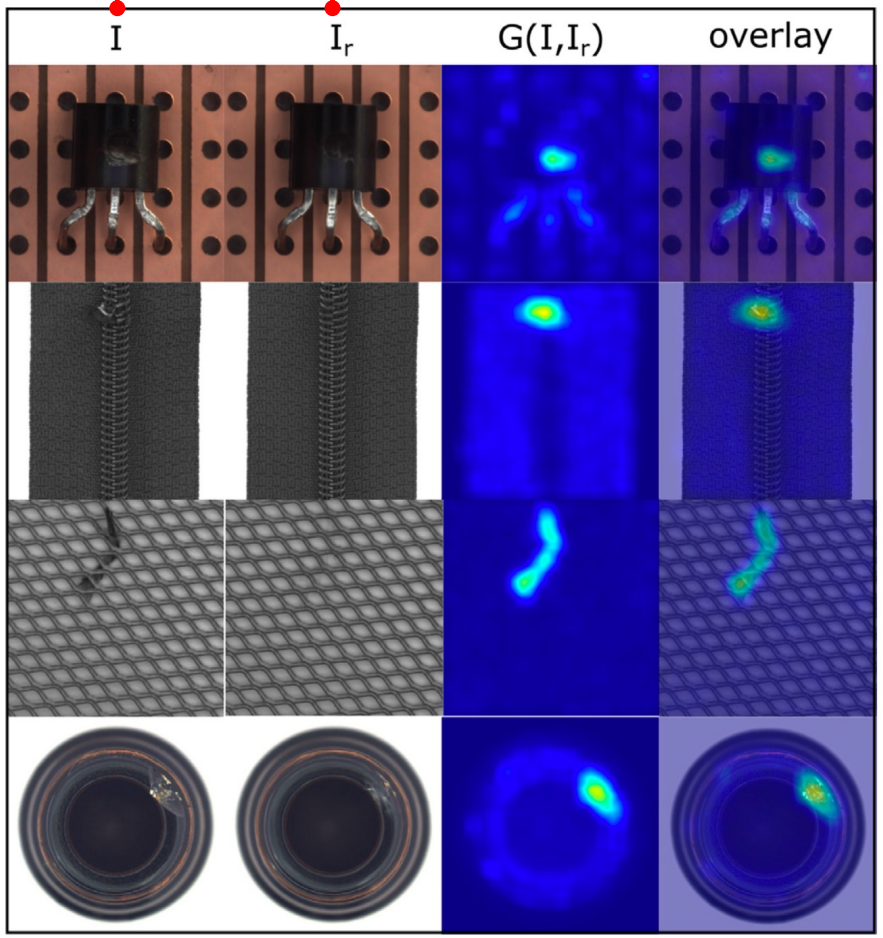
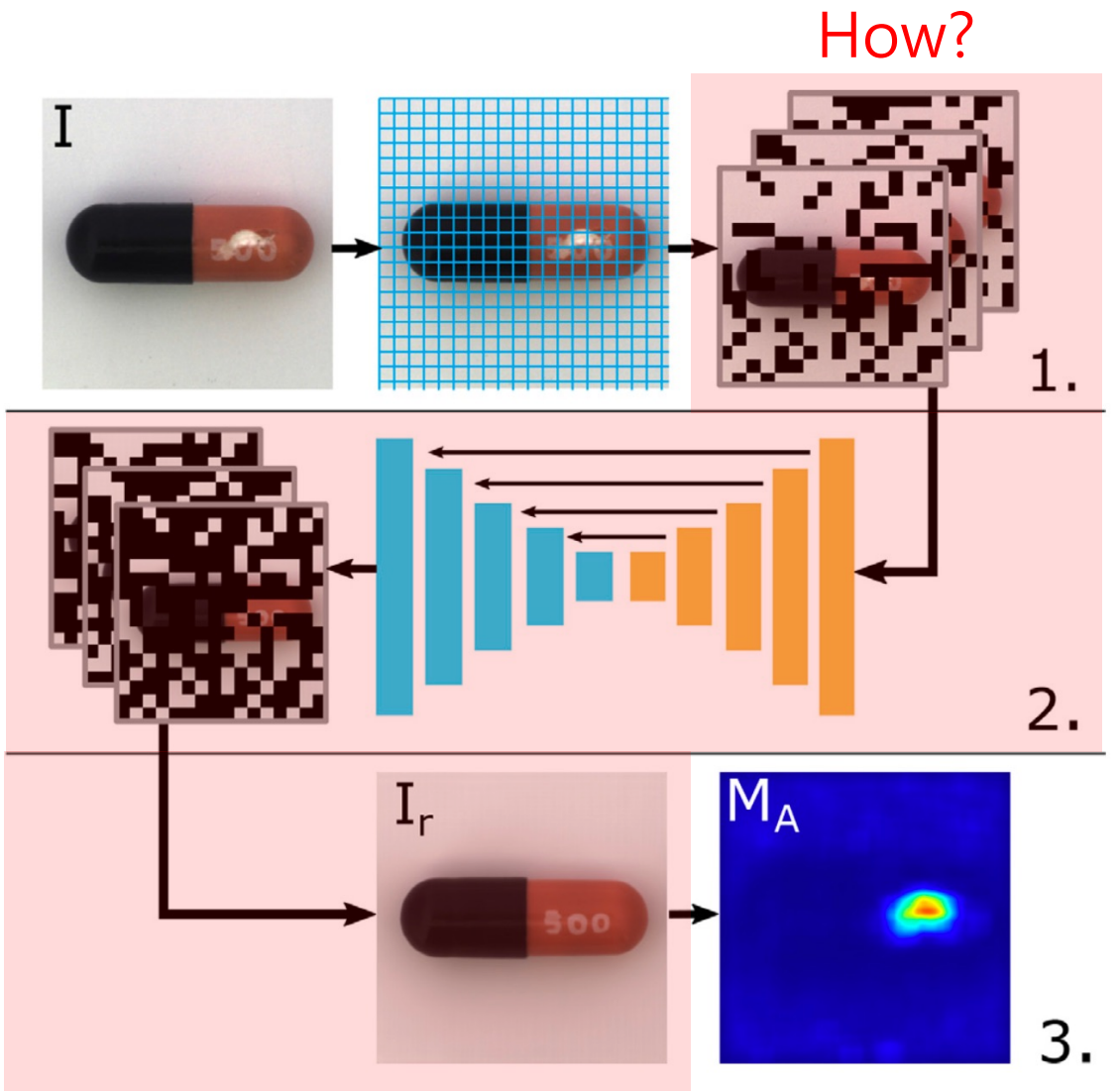


Fig. 5. Reconstruction and anomaly score estimation examples of our method. **I** is the input image, $\mathbf{I}_r$ is the fully reconstructed image and $G(\mathbf{I}, \mathbf{I}_r)$ is the MSGMS based anomaly map. The overlay of the anomaly map over the original image visualizes the localization performance of RIAD.

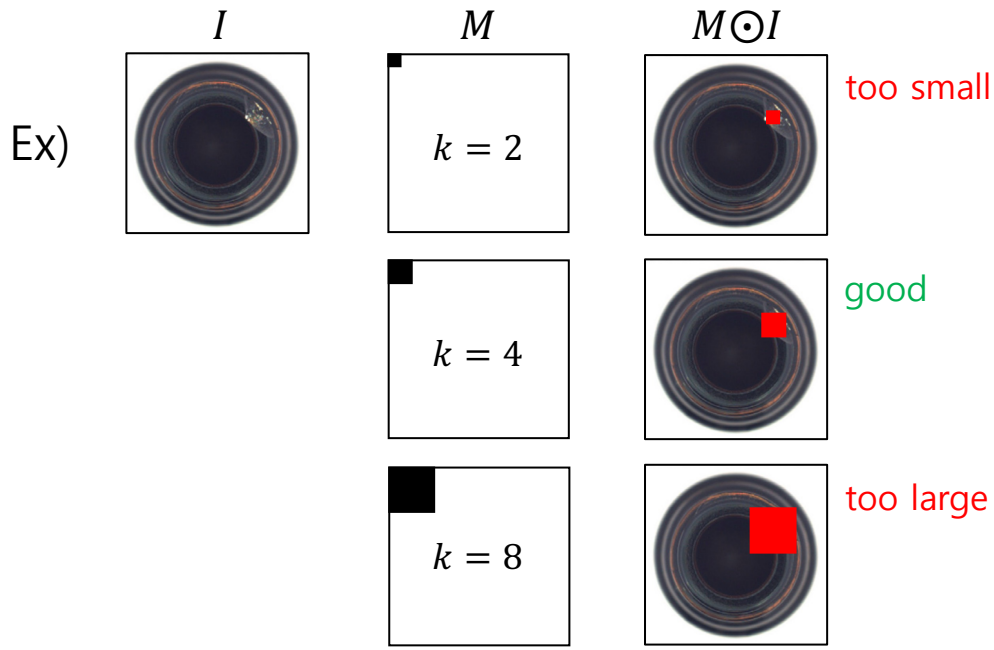
Anomaly Detection with RIAD



Training RIAD

Algorithm 2: RIAD training iteration.

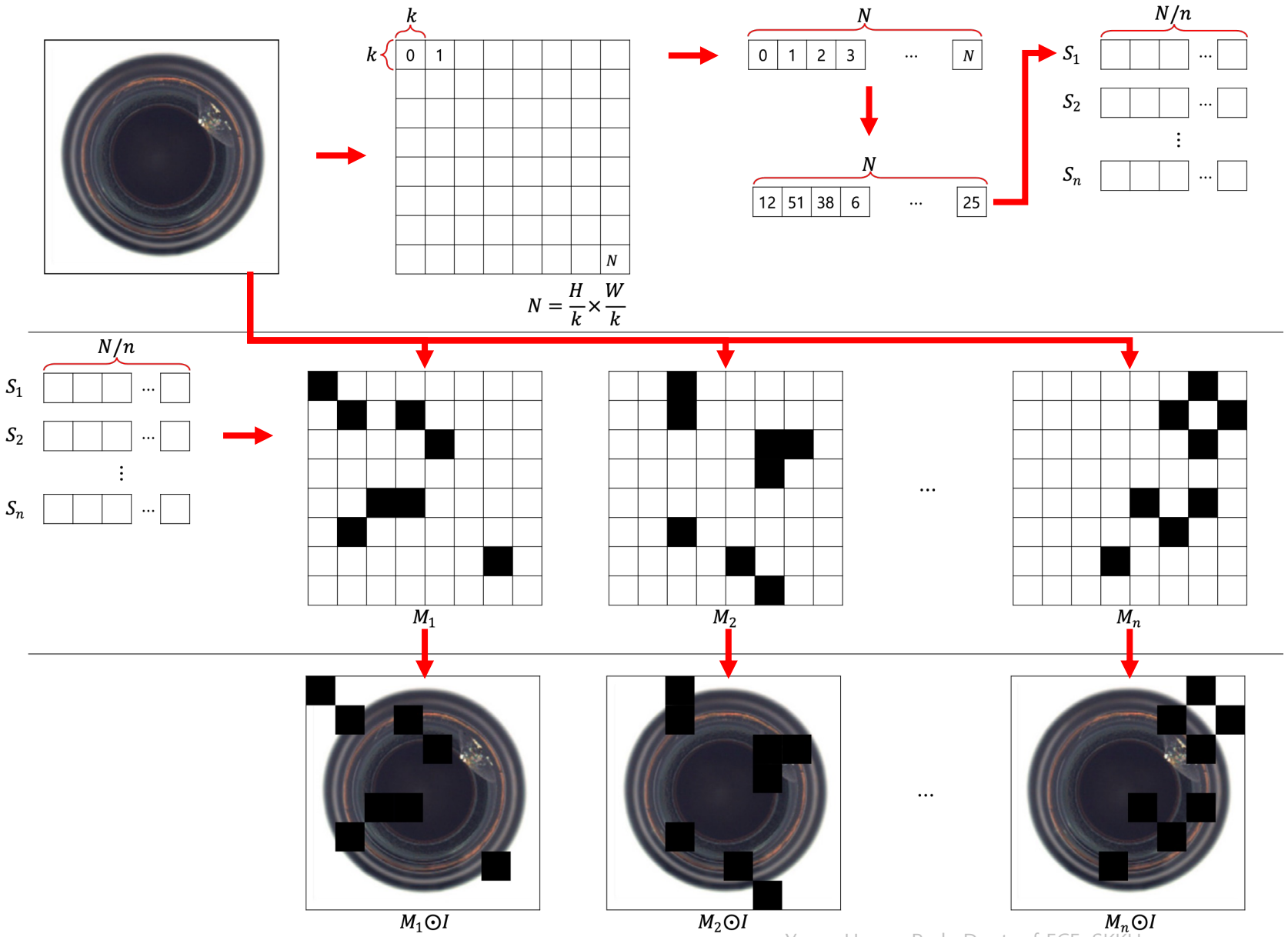
input : I ▷ input image
output: L ▷ loss for image I
K = set of region size parameters ▷ hyperparam.
k = random sample from K
I_r = reconstruction_by_inpainting (I, k)
▷ reconstruct image using Algorithm 1
$L = \lambda_G L_G(I, I_r) + \lambda_S L_S(I, I_r) + L_2(I, I_r)$ ▷ loss (7)



Algorithm 1: Reconstruction by inpainting.

input : I ▷ input image
k ▷ region size parameter
output: I_r ▷ reconstructed image
n = number of disjoint sets ▷ hyperparameter
$N = \frac{H}{k} \times \frac{W}{k}$ ▷ number of squared regions
R = permute(N) ▷ randomly permute N indices of regions of $k \times k$ pixels
$S_i = \left\{R_{i\frac{N}{n}+j}, \text{j} \in \{0, 1, \ldots, \frac{N}{n}\}\right\}$ ▷ partition R into
$S = \left\{S_i \text{ for i} \in \{0,\ldots, \text{n}\}\right\}$ ▷ n disjoint sets S_i
for S_i *in* S **do**
  $M_{S_i}^{(px)} = \begin{cases} 0, & \text{if } px \in S_i \\ 1, & \text{otherwise} \end{cases}$ ▷ binary mask, where pixels in regions S_i are set to 0
  $I_i = M_{S_i} \odot I$ ▷ mask out part of image
  I_{ri} = inpainting_model (I_i) ▷ reconstruct removed regions
end
$I_r = \sum_i^n \overline{M}_{S_i} \odot I_{ri}$ ▷ assemble full image from reconstructed regions of each I_{ri} (2)

Detail for Algorithm 1 [1/2] – preprocessing

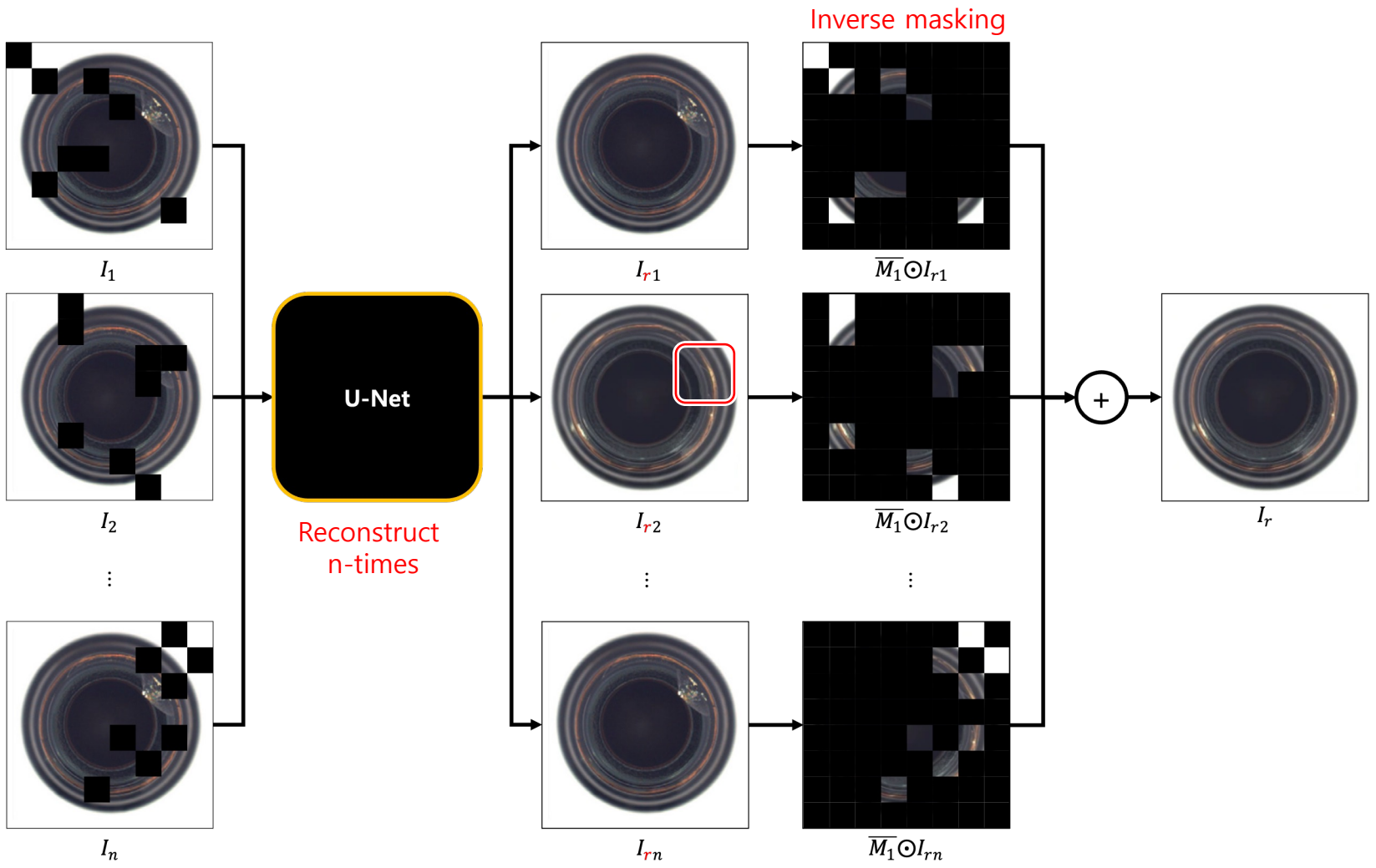


Symbol	Description
H, W	Height, width
k	region size (hyperparameter)
n	number of disjoint sets
S, S_i	Disjoint set, i-th disjoint mask

Algorithm 1: Reconstruction by inpainting.

input : I ▷ input image
k ▷ region size parameter
output: I_r ▷ reconstructed image
n = number of disjoint sets ▷ hyperparameter
$N = \frac{H}{k} \times \frac{W}{k}$ ▷ number of squared regions
R = permute(N) ▷ randomly permute N indices of regions of $k \times k$ pixels
$S_i = \left\{R_{i\frac{N}{n}+j}, \text{j} \in \{0, 1, \ldots, \frac{N}{n}\}\right\}$ ▷ partition R into
$S = \left\{S_i \text{ for i} \in \{0,\ldots, \text{n}\}\right\}$ ▷ n disjoint sets S_i
for S_i *in* S **do**
  $M_{S_i}^{(px)} = \begin{cases} 0, & \text{if } px \in S_i \\ 1, & \text{otherwise} \end{cases}$ ▷ binary mask, where pixels in regions S_i are set to 0
  $I_i = M_{S_i} \odot I$ ▷ mask out part of image
  I_{ri} = inpainting_model (I_i) ▷ reconstruct removed regions
end
$I_r = \sum_i^n \overline{M}_{S_i} \odot I_{ri}$ ▷ assemble full image from reconstructed regions of each I_{ri} (2)

Detail for Algorithm 2 [2/2] – postprocessing



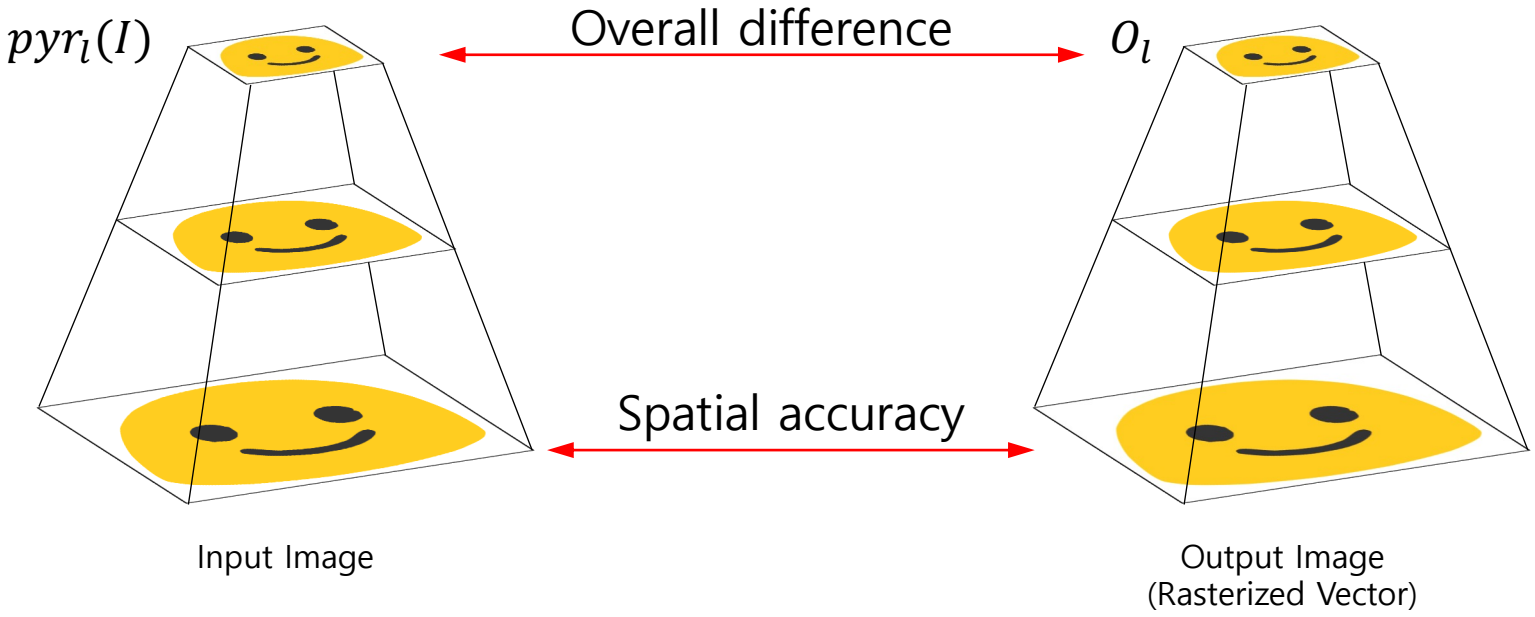
Algorithm 1: Reconstruction by inpainting.

input : I ▷ input image
k ▷ region size parameter
output: I_r ▷ reconstructed image
n = number of disjoint sets ▷ hyperparameter
$N = \frac{H}{k} \times \frac{W}{k}$ ▷ number of squared regions
R = permute(N) ▷ randomly permute N indices of regions of $k \times k$ pixels
$S_i = \left\{R_{i\frac{N}{n}+j}, \text{j} \in \{0, 1, \ldots, \frac{N}{n}\}\right\}$ ▷ partition R into
$S = \left\{S_i \text{ for i} \in \{0,\ldots, \text{n}\}\right\}$ ▷ n disjoint sets S_i
for S_i *in* S **do**
  $M_{S_i}^{(px)} = \begin{cases} 0, & \text{if } px \in S_i \\ 1, & \text{otherwise} \end{cases}$ ▷ binary mask, where pixels in regions S_i are set to 0
  $I_i = M_{S_i} \odot I$ ▷ mask out part of image
  I_{ri} = inpainting_model (I_i) ▷ reconstruct removed regions
end
$I_r = \sum_i^n \overline{M}_{S_i} \odot I_{ri}$ ▷ assemble full image from reconstructed regions of each I_{ri} (2)

Do this for multi-scale cell size (multiple k) and multi-resolution image (multiple l) !
Loop 1 Loop 2

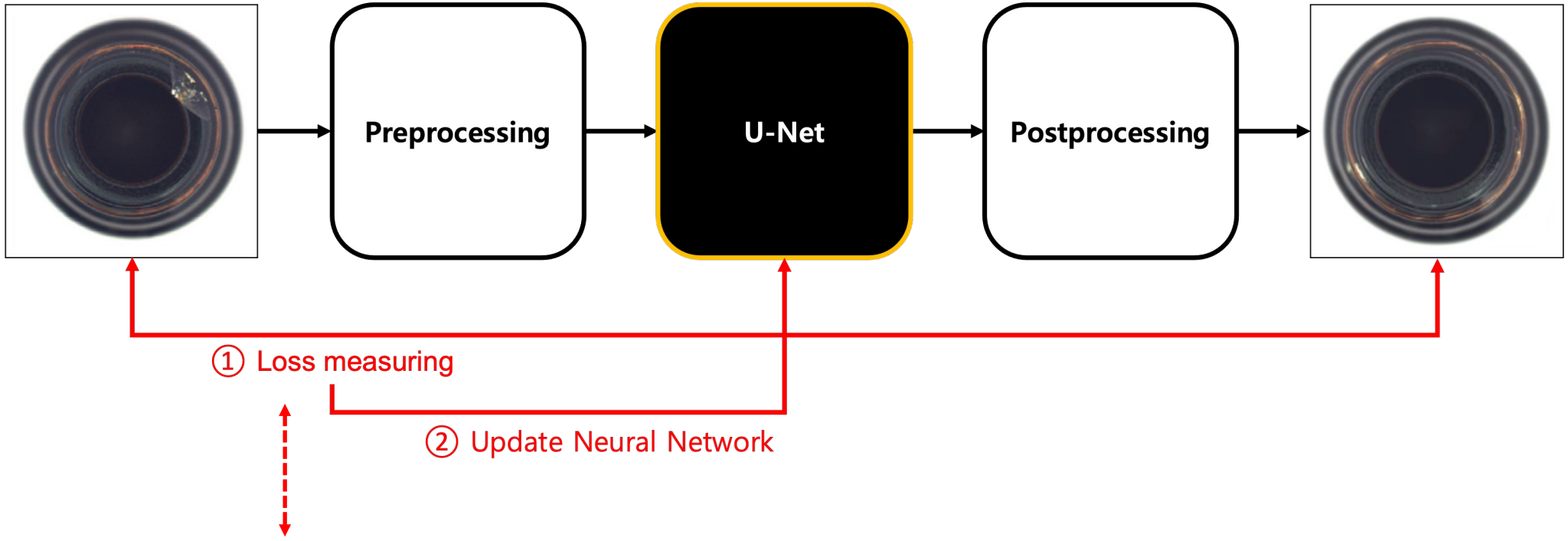
Multi-resolution Inference

Revisit Im2vec



$$\mathbb{E}_{I\sim\mathcal{D}} \sum_{l=1}^{L} \|pyr_l(I) - O_l\|^2$$

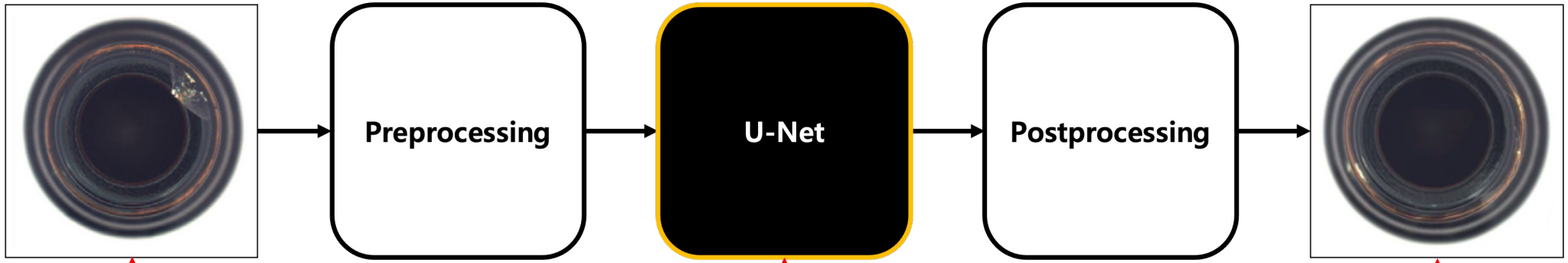
Update Neural network



$$\mathcal{L} = \lambda_G \mathcal{L}_G + \lambda_S \mathcal{L}_S + \mathcal{L}_2$$

Tracing back the loss function

Symbol	Description
L	Number of resolution
N_p	Number of pixel



① Loss measuring

② Update Neural Network

Total loss = G-loss + S-loss

$$\mathcal{L} = \lambda_G \mathcal{L}_G + \lambda_S \mathcal{L}_S + \mathcal{L}_2$$

Gradient magnitude similarity loss (cell-wise)

$$\mathcal{L}_G(I, I_r) = \frac{1}{L}\sum_{l=1}^{L}\frac{1}{N_l}\sum_{i=1}^{H_l}\sum_{j=1}^{W_l} 1 - GMS(I_l, I_{rl})_{(i,j)}$$

Structural similarity loss

$$\mathcal{L}_S(I, I_r) = \frac{1}{N_p}\sum_{i=1}^{H}\sum_{j=1}^{W} 1 - SSIM(I, I_r)_{(i,j)}$$

$$GMS(I, I_r) = \frac{2g(I)g(I_r) + c}{g(I)^2 + g(I_r)^2 + c}$$

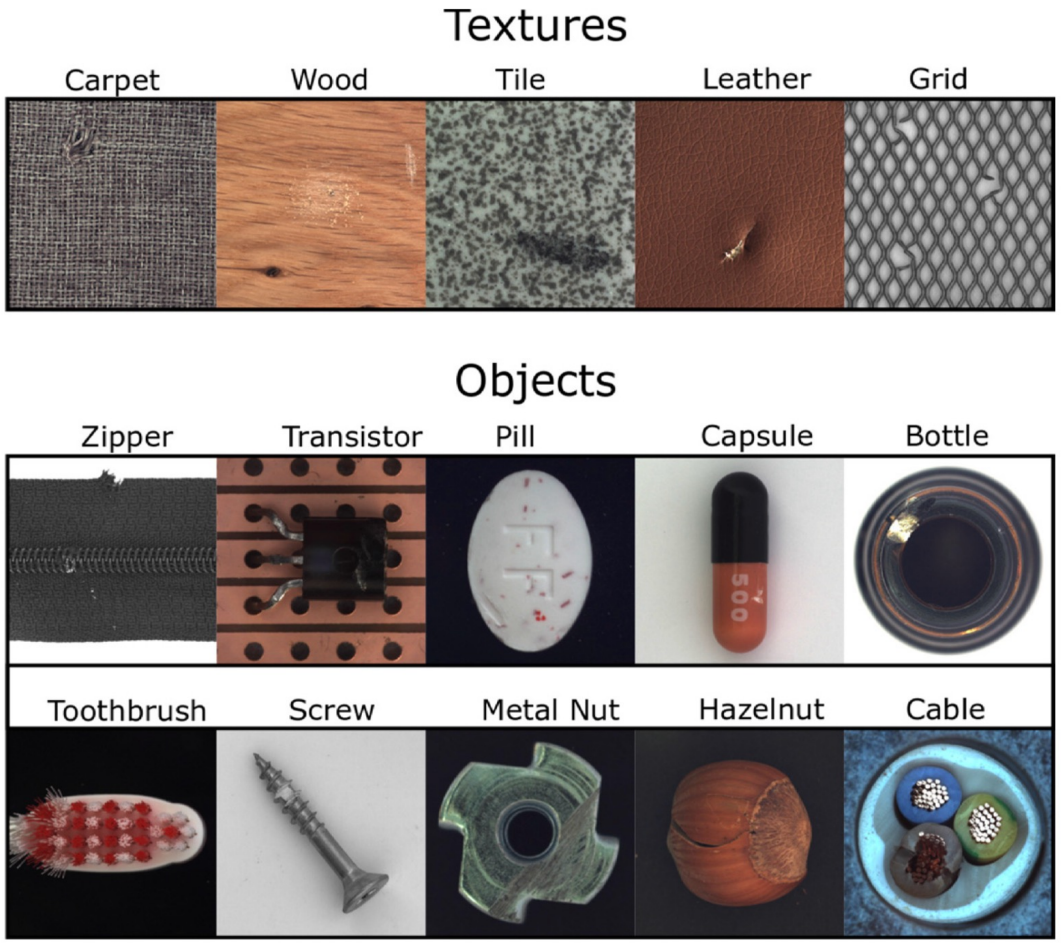
$$g(I) = \sqrt{(I * h_x)^2 + (I * h_y)^2}$$

Prewitt filters

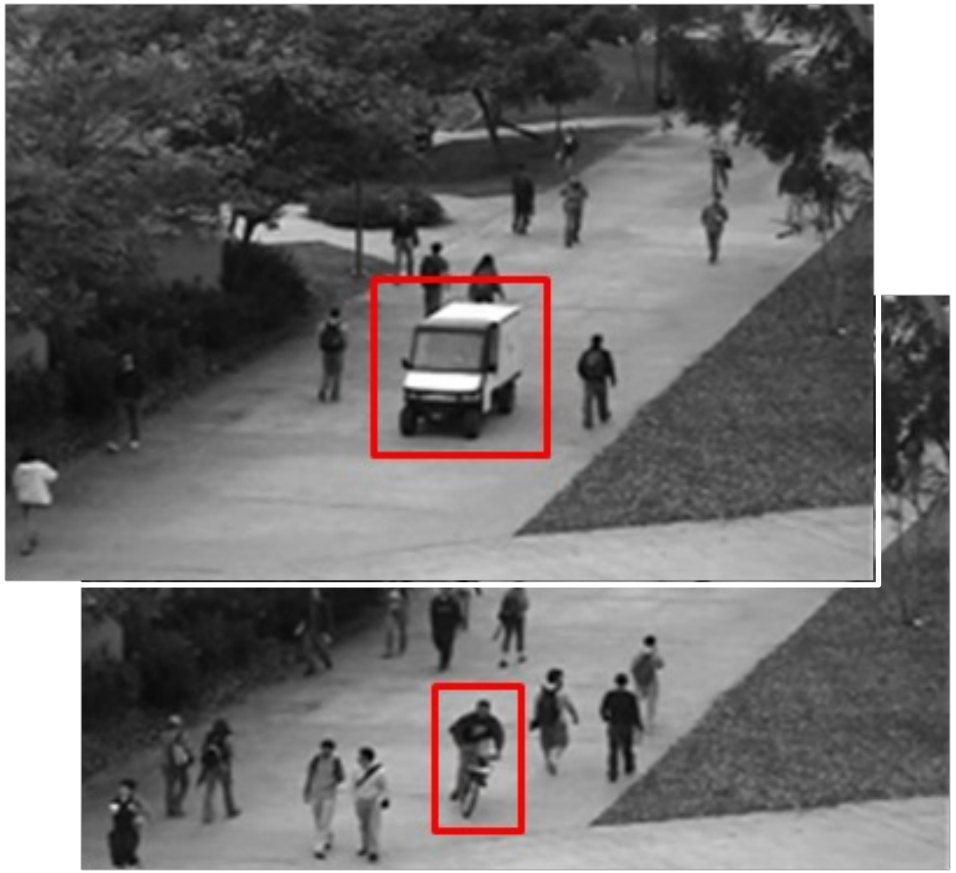
$$h_x = \begin{bmatrix} +1 & 0 & -1 \\ +1 & 0 & -1 \\ +1 & 0 & -1 \end{bmatrix} \quad h_y = \begin{bmatrix} +1 & +1 & +1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix}$$

Experiments

Dataset



One industrial anomaly detection dataset called *MVTec AD*



Two video anomaly detection datasets called *UCSD Ped2* and Avenue

Quantitative Evaluation (MVTec AD)

Table 1
Results for the task of anomaly detection on the MVTec dataset. Results are listed as ROC AUC scores and are marked individually for each class. An average score over all classes is also reported in the last row and the average score over the texture classes is computed in the row marked avg_{tex} and over the object classes in the row marked avg_{obj}.

Class	GeoTrans [6]	GANomaly [1]	ITAE [4]	US [16]	RIAD
bottle	74.4	89.2	94.1	99.0	**99.9**
capsule	67.0	73.2	68.1	86.1	**88.4**
grid	61.9	70.8	88.3	81.0	**99.6**
leather	84.1	84.2	86.2	88.2	**100**
pill	63.0	74.3	78.6	**87.9**	83.8
tile	41.7	79.4	73.5	**99.1**	98.7
transistor	86.9	79.2	84.3	81.8	**90.9**
zipper	82.0	74.5	87.6	91.9	**98.1**
cable	78.3	75.7	83.2	**86.2**	81.9
carpet	43.7	69.9	70.6	**91.6**	84.2
hazelnut	35.9	78.5	85.5	**93.1**	83.3
metal nut	81.3	70.0	66.7	82.0	**88.5**
screw	50.0	74.6	**100**	54.9	84.5
toothbrush	97.2	65.3	**100**	95.3	**100**
wood	61.1	83.4	92.3	**97.7**	93.0
avg_{tex}	58.5	76.5	82.2	91.5	**95.1**
avg_{obj}	71.6	75.4	84.8	85.8	**89.9**
avg	67.2	76.2	83.9	87.7	**91.7**

10 wins

Table 2
Localization evaluated by ROC AUC scores on the MVTec dataset. Our method ourperforms the SSIM-AE, AnoGAN, VEVAE and US methods in the overall anomaly score map ROC AUC metric.

Class	AE-SSIM [34]	AnoGAN [11]	VEVAE [35]	US [16]	RIAD
bottle	93.0	86.0	87.0	97.8	**98.4**
capsule	94.0	84.0	74.0	**96.8**	92.8
grid	94.0	58.0	73.0	89.9	**98.8**
leather	78.0	64.0	95.0	97.8	**99.4**
pill	91.0	87.0	83.0	**96.5**	95.7
tile	59.0	50.0	80.0	**92.5**	89.1
transistor	80.0	80.0	**93.0**	73.7	87.7
zipper	88.0	78.0	78.0	95.6	**97.8**
cable	82.0	78.0	90.0	**91.9**	84.2
carpet	87.0	54.0	78.0	93.5	**96.3**
hazelnut	97.0	87.0	98.0	**98.2**	96.1
metal nut	89.0	76.0	94.0	**97.2**	92.5
screw	96.0	80.0	97.0	97.4	**98.8**
toothbrush	92.0	90.0	94.0	97.9	**98.9**
wood	73.0	62.0	77.0	**92.1**	85.8
avg_{tex}	78.2	57.7	80.6	93.2	**93.9**
avg_{obj}	90.2	82.6	88.8	**94.3**	**94.3**
avg	86.2	74.3	86.1	93.9	**94.2**

9 wins

Ablation Study (MVTec AD)

Table 4
ROC-AUC anomaly detection results of our method trained and evaluated on a single value of k, where $k \in \{2, 4, 8, 16\}$. The top row shows the region size k at which the experiments were ran.

Class	2	4	8	16
bottle	99.8	99.7	**99.9**	98.8
capsule	84.2	**96.3**	94.6	**96.3**
grid	99.0	98.5	99.5	**99.7**
leather	99.0	99.9	**100**	**100**
pill	79.2	**86.2**	78.5	67.2
tile	**98.9**	92.8	75.8	65.1
transistor	83.2	84.0	91.3	**91.8**
zipper	**99.5**	98.8	97.4	98.1
cable	55.7	60.8	74.4	**87.4**
carpet	82.0	**83.5**	73.8	66.0
hazelnut	64.1	76.6	**91.0**	88.4
metal nut	72.5	62.5	**88.2**	86.3
screw	83.7	**91.1**	86.0	85.1
toothbrush	99.1	99.7	99.8	**99.9**
wood	90.9	**92.1**	86.4	84.4
avg	86.1	88.2	**89.1**	87.6

Table 5
Anomaly detection results of RIAD using various portions of masked pixels controlled by the value of $n \in \{2, 3, 4, 5\}$. The box size parameter is fixed and is $k = 8$. Results are listed as ROC-AUC scores.

Class	$n = 2$	$n = 3$	$n = 4$	$n = 5$
bottle	98.9	99.9	99.8	99.5
capsule	96.0	94.6	96.9	97.1
grid	99.5	99.5	99.3	99.5
leather	99.9	100	100	99.9
pill	71.3	78.5	77.0	77.9
tile	67.8	75.8	71.5	78.0
transistor	88.1	91.3	91.4	85.6
zipper	96.9	97.4	98.2	97.6
cable	87.4	74.4	71.5	67.5
carpet	62.7	73.8	79.0	82.7
hazelnut	73.2	91.0	86.2	85.8
metal nut	80.5	88.2	86.6	86.7
screw	87.5	86.0	89.5	90.3
toothbrush	99.9	99.8	100	100
wood	83.9	86.4	86.0	83.3
avg	86.2	89.1	88.8	88.7

* Park et al. Untitled, on going research.
** Roth, Karsten, et al. "Towards total recall in industrial anomaly detection." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2022.

Effect of GMS (MVTec AD)

Table 6
Anomaly detection results of our method with (*RIAD*) and without skip connections ($RIAD_{ns}$). $RIAD_{SSIM}$ uses *SSIM* as the anomaly score estimation method, while $RIAD_{ns}$ and *RIAD* use the *MSGMS* for anomaly score estimation. Results are listed as ROC-AUC scores.

Class	$RIAD_{ns}$	$RIAD_{SSIM}$	*RIAD*	AE_{FL}*	*PatchCore***
bottle	99.3	98.8	99.9	99.4	not provided
capsule	76.2	76.5	88.4	82.5	
grid	98.5	99.7	99.6	90.4	
leather	99.9	91.9	100	99.3	
pill	84.6	93.2	83.8	88.5	
tile	84.4	99.1	98.7	91.1	
transistor	99.4	86.6	90.9	93.2	
zipper	94.3	99.7	98.1	93.8	
cable	89.3	59.3	81.9	83.0	
carpet	62.9	86.1	84.2	85.6	
hazelnut	93.1	62.6	83.3	99.3	
metal nut	89.0	44.2	88.5	81.9	
screw	64.1	93.9	84.5	83.2	
toothbrush	99.1	96.1	100	98.6	
wood	94.3	90.1	93.0	100	
avg	88.6	85.2	91.7	91.3	99.6
wins	5	5	5		

Qualitative Evaluation (MVTec AD)

$$G(I, I_r) = 1_{H \times W} - (MSGMS(I, I_r) * f_{s_f \times s_f})$$

$$MSGMS(I, I_r) = \frac{1}{L} \sum_{l=1}^{L} GMS(I_l, I_{rl})$$

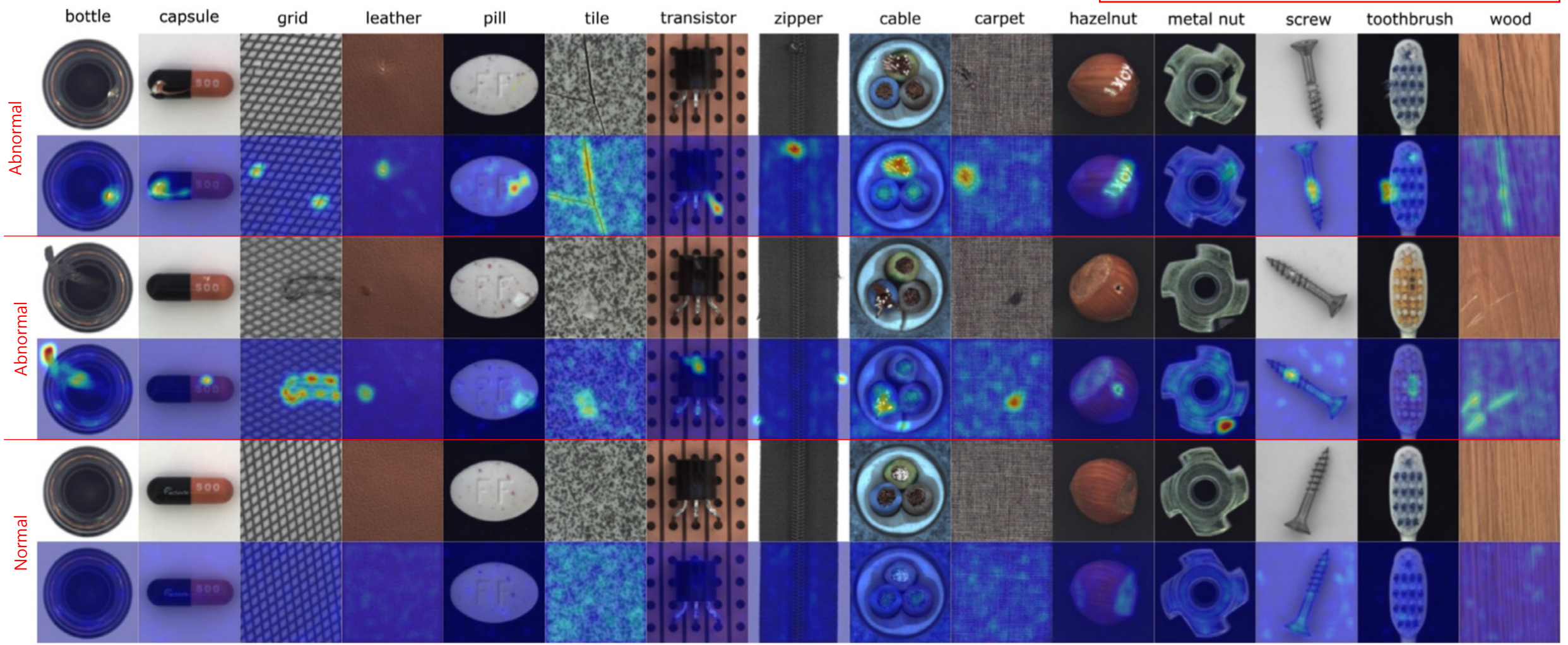


Fig. 7. Qualitative results of RIAD on the MVTec data containing anomalous images (row 1 and 3) and overlaid anomaly maps produced by RIAD (row 2 and 4). Row 5 contains non-anomalous images and row 6 contains the corresponding anomaly maps.

[33] Ionescu, Radu Tudor, et al. "Object-centric auto-encoders and dummy anomalies for abnormal event detection in video." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 2019.

Performance on Video AD (Ped2 & Avenue17)

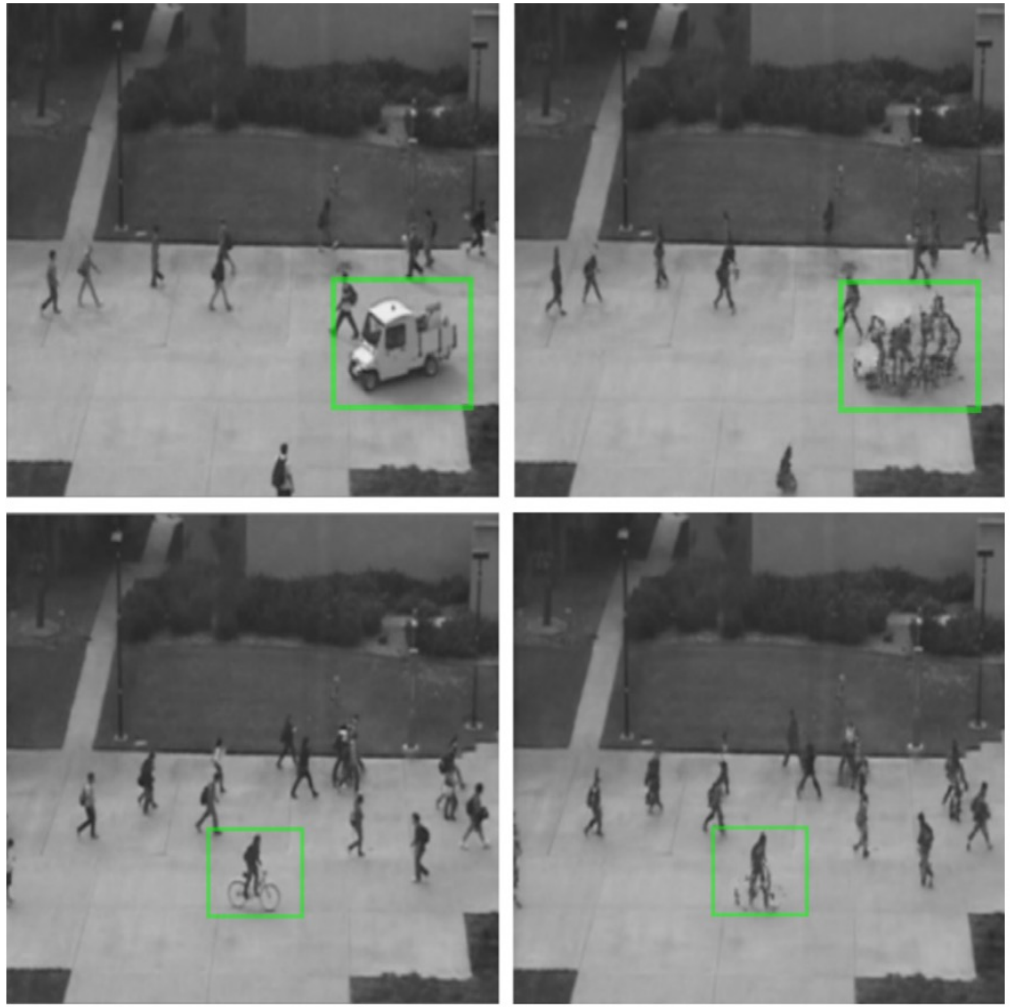


Fig. 8. Anomalies in the Ped2 data set. (Left) Original images containing anomalies. (Right) Frames reconstructed by our method. Anomalies and their reconstructions are marked in green. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3
Video anomaly detection results compared to state-of-the-art methods.

Method	Ped2	Avenue17
RIAD	92.5	88.9
Appearance Motion AE [13]	96.2	-
Future Frame Prediction [14]	95.4	-
Object Centric AE [33]	97.8	91.6
Growing Gas [37]	94.1	-
FRCN-action [32]	92.2	89.8
Conv-AE [36]	90.0	76.9

- Their method RIAD is not designed for video processing.
  - It is not a favorable setup for RIAD.
- However, RIAD shows comparative performance.

Review Comment

A few confusing or odd points are found.

- Symbol N (or n) is used for several purposes.
  - n: number of disjoint set
  - N: number of cell
  - N_p: number of pixels

- Needs word integration.
  - Multi-scale cell (for k)
  - Multi-resolution image (for l, but described as multi-scale in loss)

- The SSIM loss excludes multi-resolution manner.
  - It would be better to use multi-scale SSIM (MSSSIM) when referring to this paper.
  - However, there is no explanation for the above.

But it is overall good!

strong accept & strong recommendation

- The authors propose acceptable approaches.

- In particular, gradient magnitude similarity seems to be general-purpose as a measure.

- In the future, it would be better if the effort of hyperparameter tuning is reduced (for k and l).

Appendix A.

Knowledge Distillation-Based Anomaly Detection

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.
240 citations

Uninformed Students [1/3]

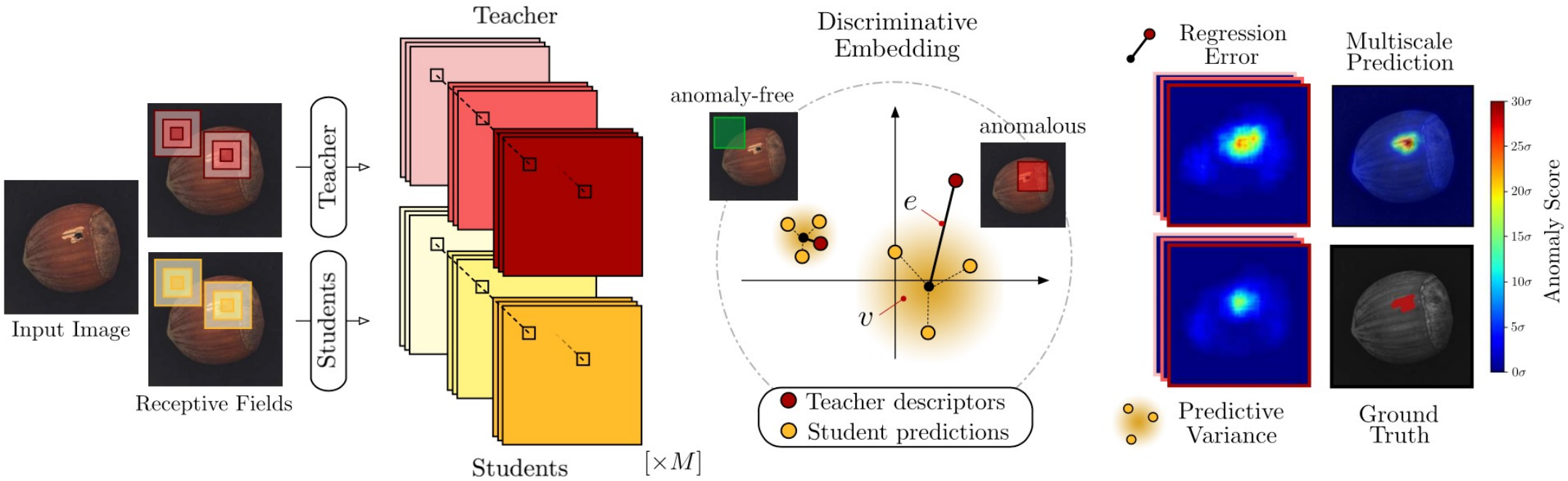


Figure 2: Schematic overview of our approach. Input images are fed through a teacher network that densely extracts features for local image regions. An ensemble of M student networks is trained to regress the output of the teacher on anomaly-free data. During inference, the students will yield increased regression errors e and predictive uncertainties v in pixels for which the receptive field covers anomalous regions. Anomaly maps generated with different receptive fields can be combined for anomaly segmentation at multiple scales.

Uninformed Students [2/3]

$$\delta^+ = ||\hat{T}(\mathbf{p}) - \hat{T}(\mathbf{p}^+)||^2$$
$$\delta^- = \min\{||\hat{T}(\mathbf{p}) - \hat{T}(\mathbf{p}^-)||^2, ||\hat{T}(\mathbf{p}^+) - \hat{T}(\mathbf{p}^-)||^2\}$$

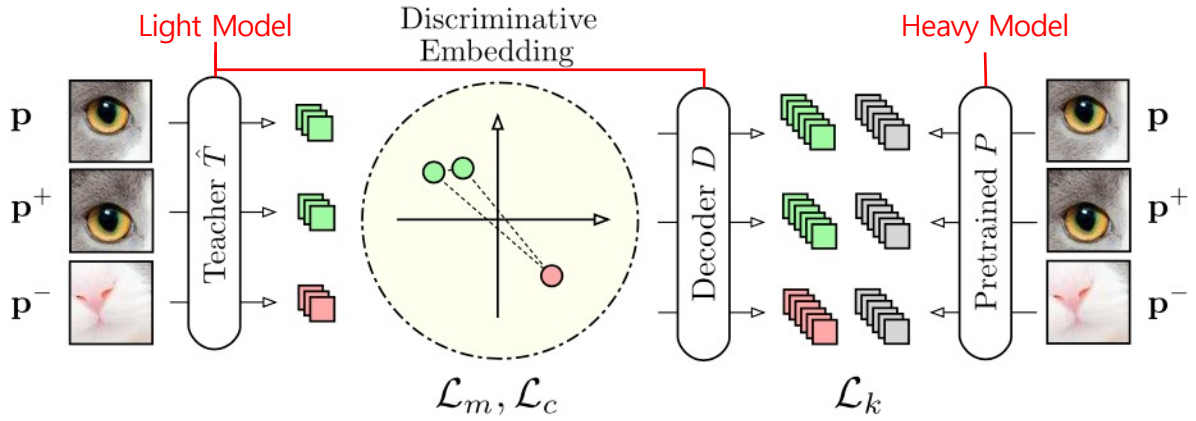


Figure 3: Pretraining of the teacher network $\hat{T}$ to output descriptive embedding vectors for patch-sized inputs. The knowledge of a powerful but computationally inefficient network P is distilled into $\hat{T}$ by decoding the latent vectors to match the descriptors of P. We also experiment with embeddings obtained using self-supervised metric learning techniques based on triplet learning. Information within each feature dimension is maximized by decorrelating the feature dimensions within a minibatch.

$$\mathcal{L}_m(\hat{T}) = \max\{0, \delta + \delta^+ - \delta^-\},$$

Metric learning
- Minimizing distance between anchor (P) and positive (P^+)
- Maximizing distance between anchor (P) and negative (P^-)

$$\mathcal{L}_c(\hat{T}) = \sum_{i \neq j} c_{ij},$$

Compactness
- Minimizing correlation between given inputs (P)
- Penalty term

$$\mathcal{L}_k(\hat{T}) = ||D(\hat{T}(\mathbf{p})) - P(\mathbf{p})||^2.$$

Knowledge Distillation
- Minimizing distance between given inputs (P)

Uninformed Students [3/3]

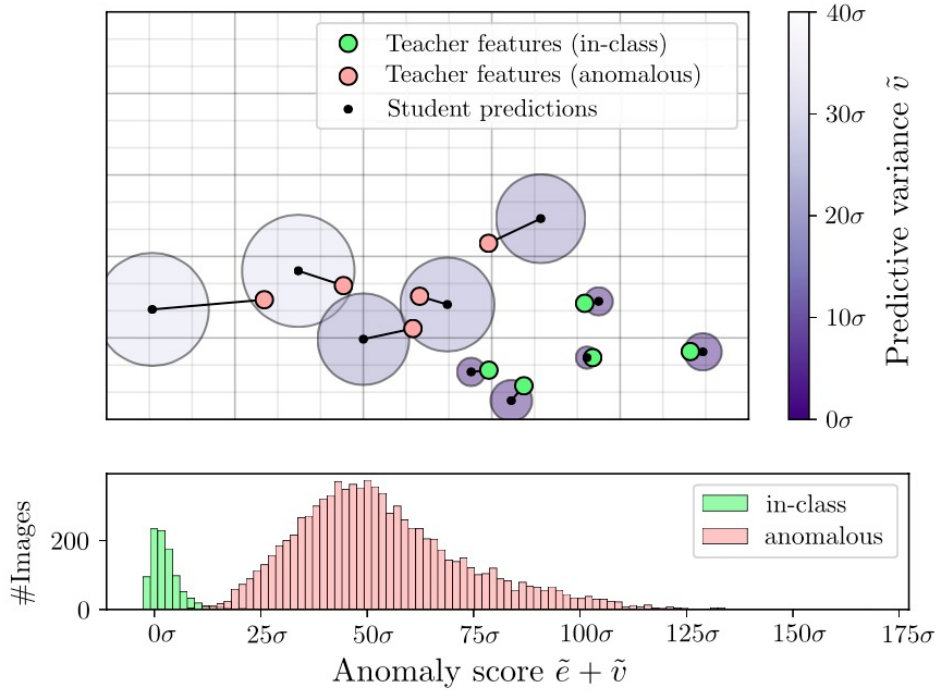


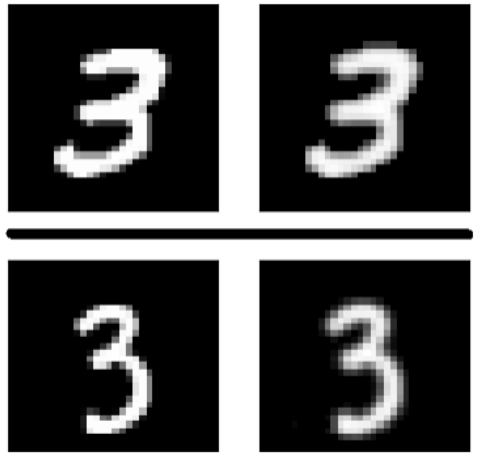
Figure 4: Embedding vectors visualized for ten samples of the MNIST dataset. Larger circles around the students' mean predictions indicate increased predictive variance. Being only trained on a single class of training images, the students manage to accurately regress the features solely for this class (green). They yield large regression errors and predictive uncertainties for images of other classes (red). Anomaly scores for the entire dataset are displayed in the bottom histogram.

Appendix B.

Geometric Transformation-Based Anomaly Detection

GeoTrans

(a) Left: original **'0'**s. Right: the **'0'**s optimized to a normality score learned on **'3'**.



(b) Left: original **'3'**s. Right: the **'3'**s optimized to a normality score learned on **'3'**.

Figure 1: Optimizing digit images to maximize the normality score

(a) and (b) represent the results of transformation attempts on abnormal to normal and normal to abnormal, respectively.