

# Supplementary Material for

## High-Fidelity Zero-Shot Texture Anomaly Localization Using Feature Correspondence Analysis

Andrei-Timotei Ardelean     Tim Weyrich  
Friedrich-Alexander-Universität Erlangen-Nürnberg  
{timotei.ardelean, tim.weyrich}@fau.de

### 1. Detailed Quantitative Comparison

In Table 1 we present the detailed breakdown of the metrics into texture classes for all datasets. We report the PRO(0.3), AUROC, and  $F_1$  metrics as introduced in the main paper.

### 2. Additional Qualitative Comparison

As mentioned in the main paper, we add here more visualizations for an extensive quality comparison between our method and the current state-of-the-art zero-shot method of Aota *et al.* [1].

#### 2.1. Thresholded Results

We visually compare the anomalous regions extracted by different methods, by thresholding the anomaly maps. We compute the optimal threshold with respect to the  $F_1$  measure for each texture class individually and display the anomalous area in the original texture. This visualization is included in Figure 1.

#### 2.2. Complete Results

To facilitate the comparison between our method and the main baseline of Aota *et al.* [1], we present the entirety of our anomaly maps using a static html tree. The results can be conveniently browsed under [reality.tf.fau.de/pub/ardelean2024highfidelity.html](http://reality.tf.fau.de/pub/ardelean2024highfidelity.html).

To get from continuous anomaly scores to the visualizations we perform the following. We consider each texture category/class independently and normalize the scores (same linear scale and bias within each category) for each anomaly prediction, mapping the whole class’s minimum and maximum to 0 and 1, respectively. As described in the main paper, the borders ( $\approx 10\%$ ) of each image are ignored during normalization, and they are afterward clipped accordingly. The continuous scores are then mapped to the “Reds” color scheme.

### 3. Failure Cases

Figure 2 investigates the failure cases of our method. FCA is designed to perform anomaly localization for textured images by comparing patches with a global reference. As the complexity of the normal texture distribution increases, a single aggregated reference cannot fully capture this complexity. As shown in the first two rows, complex bimodal textures with large periodicity present difficulties for our method when a single reference is used. Taking the  $k$ -NN references alleviates the issue; however, as explained in the main paper, this solution is not feasible for high-resolution images. The third row in Figure 2 shows a different type of failure. Our formulation assumes that there are enough normal patches in the original image to infer normalcy. When testing on a center-cropped image from MVTEC AD tile class (center crop of  $512 \times 512$ ), one can see that if anomalies cover the majority of the image, our method is unable to properly detect these defects.

### 4. Additional Study

We further inspect the proposed approach through a direct ablation and study of the FCA sensitivity to its parameters.

#### 4.1. Direct Ablation

In Table 2, we perform a direct ablation of the components which constitute the proposed method. Ablating **F** means instead of the default features, extracted by a WideResnet-50 [14] pretrained on ImageNet, we use the same architecture but no pretraining, *i.e.* randomly initialized weights. Ablating **R** means instead of our Wasserstein-optimal feature set (see Section 3.3 main text), we simply take the average of all feature patches in the image. Ablating **HR** gives the difference between running at full resolution ( $1024 \times 1024$ ) over low resolution ( $320 \times 320$ ). Ablation of **FCA** is performed by replacing our novel Patch Statistics Comparison method with the EMD between histograms of the features in each patch. In this case, the reference is the global histogram.

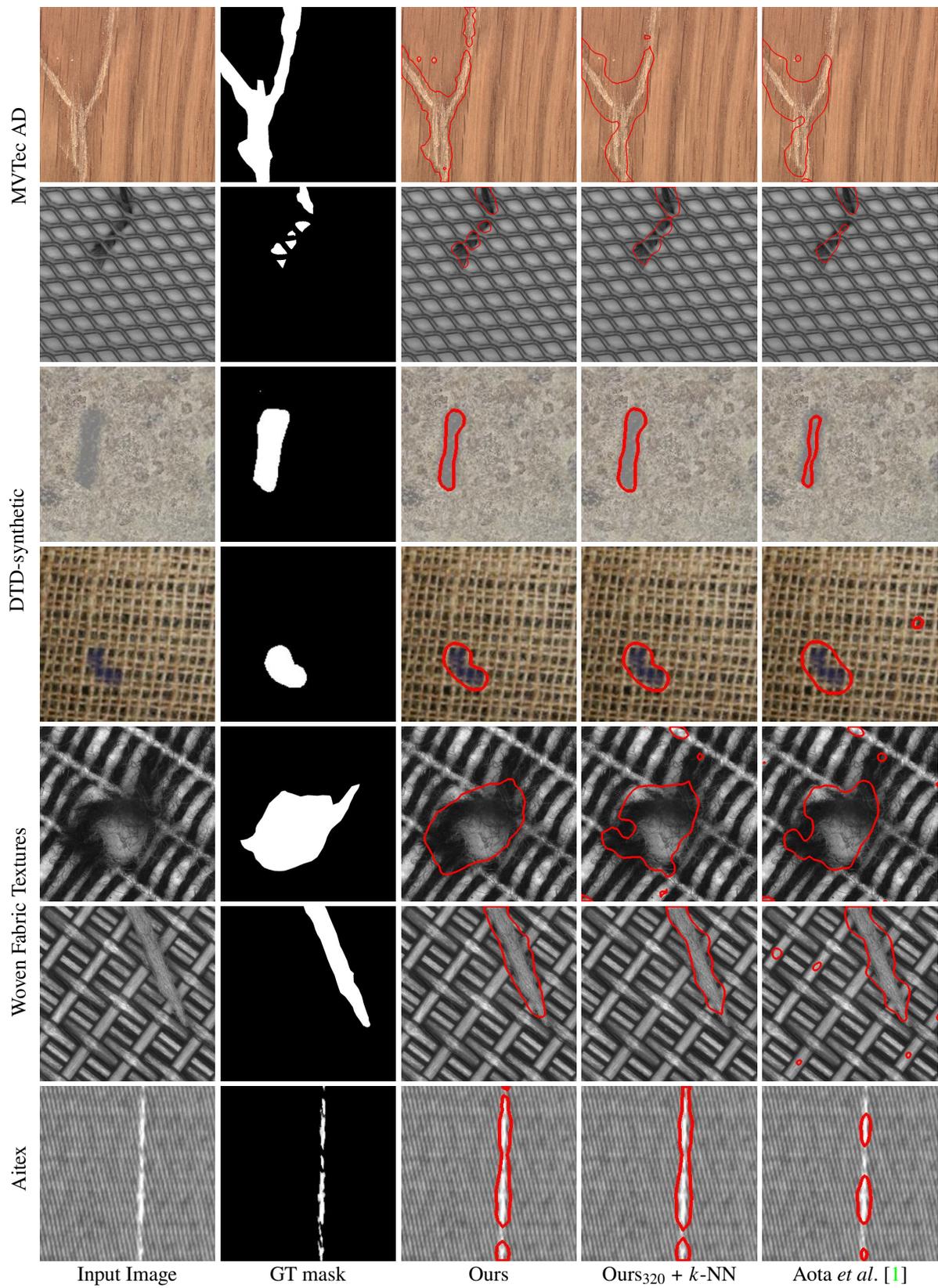


Figure 1. Qualitative comparison on challenging examples, displaying anomalous regions detected by each method via thresholding of the predicted anomaly maps. The respective thresholds are chosen to be  $F_1$ -optimal. The thresholded (binary) maps are represented through their enclosing contours; all images are shown after cropping to the center.

	Ours			Ours + KNN			Aota <i>et al.</i> [1]		
	PRO	AUROC	F1	PRO	AUROC	F1	PRO	AUROC	F1
<b>MVTec AD [4]</b>									
carpet	95.44	98.30	<b>72.58</b>	<b>96.92</b>	98.81	71.53	96.13	<b>98.83</b>	69.97
grid	<b>98.07</b>	<b>99.46</b>	<b>61.62</b>	97.77	99.27	52.41	97.01	99.12	51.89
leather	98.90	99.45	<b>66.06</b>	<b>98.92</b>	<b>99.52</b>	60.26	98.13	99.47	58.74
tile	<b>96.33</b>	<b>98.22</b>	<b>82.16</b>	88.95	94.31	65.55	84.29	93.41	63.74
wood	<b>97.18</b>	<b>98.22</b>	<b>76.34</b>	95.22	96.98	62.42	93.56	96.54	58.68
<b>DTD-Synthetic [1]</b>									
Blotchy_099	98.67	99.55	78.92	<b>98.73</b>	<b>99.57</b>	<b>79.50</b>	97.60	99.19	69.56
Marbled_078	97.97	99.33	76.25	<b>98.05</b>	<b>99.37</b>	<b>76.60</b>	96.51	98.80	66.11
Mesh_114	94.44	97.63	65.46	<b>95.91</b>	<b>98.16</b>	<b>66.95</b>	95.14	97.75	64.16
Stratified_154	98.78	99.15	66.67	<b>98.81</b>	99.19	<b>66.80</b>	98.53	<b>99.25</b>	64.48
Woven_068	97.24	98.86	70.51	<b>97.31</b>	<b>98.92</b>	<b>70.72</b>	95.34	98.29	65.61
Woven_125	98.51	99.52	77.03	<b>98.66</b>	<b>99.56</b>	<b>77.59</b>	97.00	98.99	67.50
Fibrous_183	96.95	98.96	72.82	<b>97.21</b>	<b>99.06</b>	<b>73.29</b>	94.42	98.20	65.42
Matted_069	<b>89.43</b>	99.33	76.15	88.89	<b>99.37</b>	<b>76.60</b>	89.34	99.17	68.64
Perforated_037	94.55	96.60	64.66	<b>96.62</b>	<b>97.76</b>	<b>68.78</b>	95.74	97.05	67.41
Woven_001	94.73	98.93	66.17	<b>97.79</b>	<b>99.59</b>	<b>68.96</b>	96.51	99.42	63.59
Woven_104	89.98	96.84	64.93	<b>90.40</b>	<b>96.96</b>	<b>66.05</b>	89.78	96.60	65.28
Woven_127	85.26	91.63	58.88	<b>92.82</b>	<b>94.62</b>	<b>69.58</b>	85.96	92.22	63.81
<b>WFT [5]</b>									
texture_1	<b>92.38</b>	<b>97.98</b>	<b>80.34</b>	85.36	94.25	68.46	89.01	95.94	73.12
texture_2	86.77	<b>98.56</b>	<b>77.91</b>	<b>87.12</b>	98.13	77.05	80.16	96.28	71.02
<b>Aitex [12]</b>									
t_00	<b>78.00</b>	<b>95.08</b>	49.41	75.19	94.08	<b>49.56</b>	62.91	90.21	45.33
t_01	76.85	91.28	69.62	<b>80.08</b>	<b>92.34</b>	<b>73.47</b>	70.71	91.31	71.04
t_02	96.34	<b>99.32</b>	56.30	<b>96.57</b>	<b>99.33</b>	<b>56.93</b>	92.92	99.02	53.45
t_03	88.77	97.94	<b>68.04</b>	<b>89.44</b>	<b>97.97</b>	67.75	87.43	97.59	65.83
t_04	<b>99.39</b>	<b>99.79</b>	<b>72.44</b>	<b>99.39</b>	<b>99.79</b>	72.23	97.89	99.75	72.04
t_05	<b>98.15</b>	<b>99.17</b>	<b>49.34</b>	98.04	99.11	47.66	97.95	99.05	43.41
t_06	<b>99.97</b>	<b>99.99</b>	71.60	99.97	<b>99.99</b>	70.73	<b>99.98</b>	<b>99.99</b>	<b>76.54</b>

Table 1. Metrics breakdown into texture classes.

F	R	HR	FCA	PRO(0.3)	AUROC
✗	✓	✓	✓	70.24	87.99
✓	✗	✓	✓	69.62	78.58
✓	✓	✗	✓	95.46	97.74
✓	-	✓	✗	95.35	98.00
✓	✓	✓	✓	<b>97.18</b>	<b>98.73</b>

Table 2. Ablating different components of our pipeline: F - Feature Extractor, R - Reference, HR - Running the method at high resolution. See text for more details.

## 4.2. Sensitivity to parameters

Our method has only 3 parameters:  $T$ ,  $\sigma_p$ , and  $\sigma_s$ . Additionally, the image size can be considered a preprocessing parameter that influences the performance of the method. In the main paper, we demonstrated the robustness of our approach to the choice of these parameters by running all experiments with fixed  $\sigma_p = 3$  and  $\sigma_s = 1$ . We set  $T = 9$

when running at full resolution and  $T = 3$  at low resolution ( $320 \times 320$ ). This fixed setup performed well on all datasets tested. In Figure 3 we further show the relationship between varying parameters and the performance on MVTEC AD [3]. We run our method at full resolution ( $1024 \times 1024$ ) in three settings with  $T \in \{7, 9, 11\}$ . In each case, we vary  $\sigma_p$  and  $\sigma_s$  and observe that FCA is robust to these variations. The performance deteriorates significantly when  $T = 7$ , which is primarily due to a reduced receptive field leading to more false positive predictions. In general,  $T$  should be set large enough to capture the periodicity of the texture.

## 5. Miscellaneous Details

When evaluating the different options for patch statistics comparison (Section 3.2 main text) we use  $\sigma = 6$  for Gaussian spatial smoothing for all methods. In the case of FCA,  $\sigma_p = 6$  and  $\sigma_s = 3$ . These hyperparameters are calibrated for the preliminary experiment, where the size of feature maps is

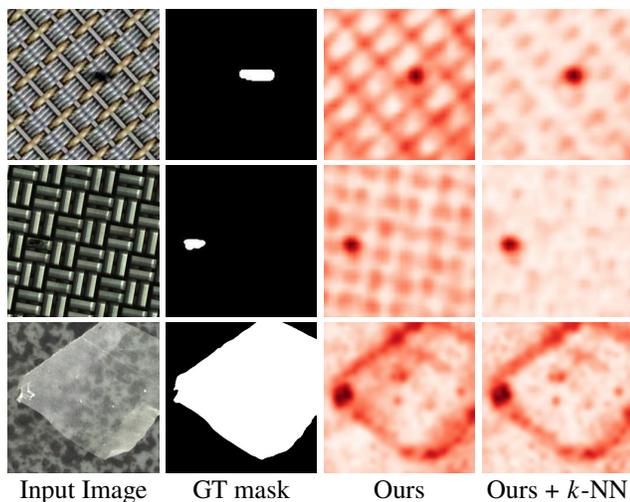


Figure 2. Visualization of failure cases caused by complex normal distributions with large texture period or anomaly regions that cover the majority of an image. All images are shown after cropping to the center.

significantly larger ( $256 \times 256$ ).

When computing the histogram method for comparing patches we scale the features to the interval  $[0.0, 1.0]$  and use a fixed number of 10 bins. Using more bins yields marginal improvements in quality but increased computational cost. We also scale features to  $[0.0, 1.0]$  when computing SWW and FCA to ensure a similar scale across channels and images.

When computing the scores for Aota *et al.* [1], SAA+ [6], and April-GAN [7] the official code has been used and run with the default parameters; for Saliency [8] we used a public unofficial implementation<sup>1</sup>; for Bellini *et al.* [2] we reimplemented the method described by the authors; for PatchCore [10], RD++ [13], MAEDAY [11], and WinCLIP [9] we take the results directly from the papers which introduce the respective methods.

## 6. Code

The source code for the introduced method is available at [github.com/TARdelean/AnomalyLocalizationFCA](https://github.com/TARdelean/AnomalyLocalizationFCA).

## References

- [1] Toshimichi Aota, Lloyd Teh Tzer Tong, and Takayuki Okatani. Zero-shot versus many-shot: Unsupervised texture anomaly detection. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5564–5572, 2023. 1, 2, 3, 4
- [2] Rachele Bellini, Yanir Kleiman, and Daniel Cohen-Or. Time-varying weathering in texture space. *ACM Transactions on Graphics (TOG)*, 35(4):1–11, 2016. 4
- [3] Paul Bergmann, Kilian Batzner, Michael Fauser, David Sattlegger, and Carsten Steger. The mvtec anomaly detection dataset: a comprehensive real-world dataset for unsupervised anomaly detection. *International Journal of Computer Vision*, 129(4):1038–1059, 2021. 3
- [4] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9592–9600, 2019. 3
- [5] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In *VISIGRAPP*, 2018. 3
- [6] Yunkang Cao, Xiaohao Xu, Chen Sun, Yuqi Cheng, Zongwei Du, Liang Gao, and Weiming Shen. Segment any anomaly without training via hybrid prompt regularization. *arXiv preprint arXiv:2305.10724*, 2023. 4
- [7] Xuhai Chen, Yue Han, and Jiangning Zhang. A zero-/few-shot anomaly classification and segmentation method for cvpr 2023 vand workshop challenge tracks 1&2: 1st place on zero-shot ad and 4th place on few-shot ad. *arXiv preprint arXiv:2305.17382*, 2023. 4
- [8] Ming-Ming Cheng, Niloy J Mitra, Xiaolei Huang, Philip HS Torr, and Shi-Min Hu. Global contrast based salient region detection. *IEEE transactions on pattern analysis and machine intelligence*, 37(3):569–582, 2014. 4
- [9] Jongheon Jeong, Yang Zou, Taewan Kim, Dongqing Zhang, Avinash Ravichandran, and Onkar Dabeer. Winclip: Zero-/few-shot anomaly classification and segmentation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19606–19616, June 2023. 4
- [10] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14318–14328, 2022. 4
- [11] Eli Schwartz, Assaf Arbelle, Leonid Karlinsky, Sivan Harary, Florian Scheidegger, Sivan Doveh, and Raja Giryes. MAEDAY: MAE for few and zero shot Anomaly-detection. *arXiv preprint arXiv:2211.14307*, 2022. 4
- [12] Javier Silvestre-Blanes, Teresa Alberro-Albero, Ignacio Miralles, Rubén Pérez-Llorens, and Jorge Moreno. A public fabric database for defect detection methods and results. *Autex Research Journal*, 19(4):363–374, 2019. 3
- [13] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan T.M. Duong, Chanh D. Tr. Nguyen, and Steven Q. H. Truong. Revisiting reverse distillation for anomaly detection. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24511–24520, June 2023. 4
- [14] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. In Edwin R. Hancock Richard C. Wilson and William A. P. Smith, editors, *Proceedings of the British Machine Vision Conference (BMVC)*, pages 87.1–87.12. BMVA Press, September 2016. 1

<sup>1</sup><https://github.com/congve1/SaliencyRC>

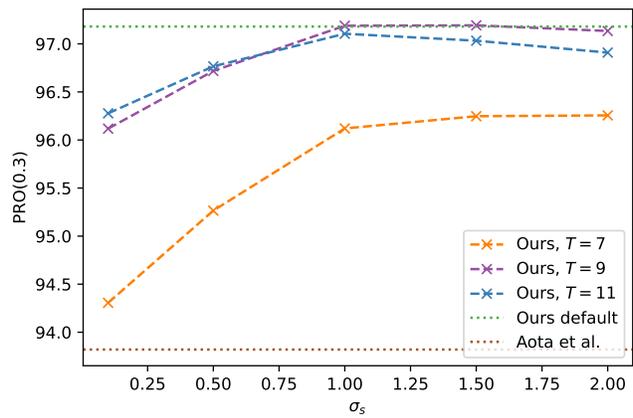
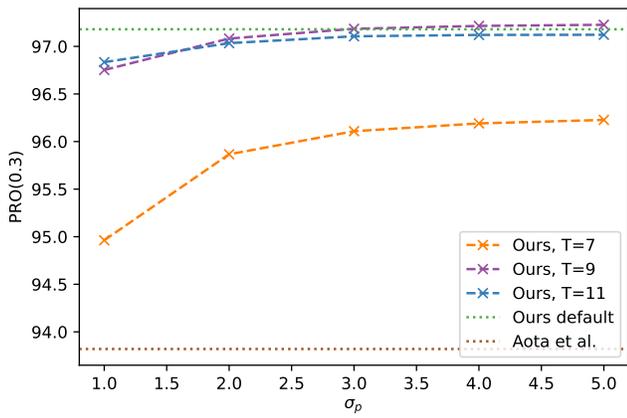


Figure 3. Sensitivity of the method to parameters' variations for the MVTEC AD dataset, at full resolution ( $1024 \times 1024$ ). 'Ours default' is obtained with  $T = 9$ ,  $\sigma_p = 3.0$ , and  $\sigma_s = 1.0$ .