UniNet: A Contrastive Learning-guided Unified Framework with Feature Selection for Anomaly Detection

Supplementary Material

1. Overview

The supplementary material is organized as follows: Appendix 2 provides additional details on the datasets and the implementation of UniNet. Appendix 3 presents further experimental results on the MVTec AD, BTAD, VisA, and MVTec 3D-AD datasets, as well as the complexity analysis of UniNet. Appendix 4 presents supplementary ablation study results. Appendix 5 includes additional visualization results across various datasets. Appendix 6 discusses the limitations and potential directions for future work.

2. Experimental setup

2.1. Datasets

MVTec AD [2] is a widely used dataset for industrial anomaly detection, comprising 15 object and texture categories with a total of over 5,000 images. The dataset contains various types of anomalies, such as scratches and crackes. Each category includes a training set consisting solely of normal images and a test set of containing both normal and abnormal images.

BTAD [24] is a real-world industrial dataset consisting of 3 different types of industrial products. The dataset contains 2830 images, with 400, 1,000, and 399 training images in class 1, 2, and 3, respectively.

VisA [36] is a publicly available dataset for visual anomaly detection, comprising 12 categories and a total of 10,821 high-resolution images from diverse domains such as electronics, food, and industrial parts. The dataset includes both normal and anomalous samples, with detailed annotations for the anomalies.

MVTec 3D-AD [3] is a multi-modal dataset that includes two different modality: RGB images and Point Clouds. The dataset consists of 10 real-world categories with a total of 4147 high-resolution images.

VAD [1] is a newly introduced supervised benchmark designed to encompass a wider array of complex anomalies and substantial intra-class variability in anomalous-free images. The dataset contains 5,000 object images, with 165 unseen anomalous images reserved for testing.

APTOS [27] is a collection of color fundus images from the 2019 APTOS blindness detection challenge. Each image is associated with a label (ranging from 0 to 4) that indicates the severity of diabetes retinopathy, with grade 0 representing normal images.

OCT2017 [16] is a dataset of optical coherence tomography images, with one class labeled as normal and three other classes labeled as abnormal. The dataset contains over 20,000 images, with 1000 images used for testing.

Algorithm 1: Weighted Decision Mechanism for anomaly detection

Input: Pixel-level anomaly map M_{AS} for each sample and outputs F_S^i and F_T^i from the S-T models

Output: Anomaly score S_{AD} for each sample

Function `Main`

  `// Low similarity value` v_l `generation`

  for $i = 1, 2, \ldots, n$ **do**

    minimize vector-wise cosine similarity between $\{F_S^i, F_T^i\}$ by $d(\cdot)$ and obtain v_{l_i} by $max(\cdot)$

  `// Transform into the probability distribution`

  $v_p = \text{Softmax}(v_l)$

  `// Weight` v_w `generation`

  for $i = 1, 2, \ldots, 2n$ **do**

    if $v_{p_i} > \frac{1}{2n} \sum_{i=1}^{2n} v_{p_i}$ **then**

      Incorporate large values into $\mathcal{P}$

  Compute v_w using $\mathcal{P}$ based on Eq. (14)

  `// Evaluate anomaly score` S_{AD}

  $S_{AD} = \frac{1}{v_w} \sum_{1}^{K=v_w} \text{top}K(M_{AS})$

  return S_{AD}

ISIC2018 [7] is a collection of skin disease images from Task 3 of the ISIC2018 challenge. The dataset includes seven classes, with *nevus* labeled as the normal class and the remaining classes representing various types of anomalies. Following [10], 6705 normal images from training set are used, while the validation set of 193 images serves as the test set.

Kvasir [15], **CVC-ClinicDB** [4], and **CVC-ColonDB** [28] are three polyp segmentation datasets, containing a total of 1,000, 612, and 379 images, respectively, sourced from diverse imaging clinics and centers. Each image is accompanied by a corresponding pixel-level mask.

Ped2 [19] is a dataset designed for video anomaly detection, consisting of 2.6K frames for training and 2.0K frames for testing. The anomalies in the dataset include cycling, skateboarding, *etc*.

(a) MVTec AD

Category	**UniNet(Ours)**	ReContrast [10]	RealNet [35]	ReConPatch [14]	GLAD [32]	RD++ [29]	DMAD [21]
Carpet	100.0 / 99.2 / 97.5	99.8 / 99.3 / 97.9	99.8 / 99.2 / 96.4	99.6 / 98.8 / –	99.0 / 98.5 / –	100.0 / 99.2 / 97.7	– / – / –
Grid	99.5 / 99.4 / 98.0	100.0 / 99.2 / 97.8	100.0 / 99.5 / 97.3	100.0 / 99.0 / –	100.0 / 99.6 / –	100.0 / 99.3 / 97.7	– / – / –
Leather	100.0 / 99.5 / 98.3	100.0 / 99.5 / 99.2	100.0 / 99.8 / 96.2	100.0 / 96.0 / –	100.0 / 99.8 / –	100.0 / 99.4 / 99.2	– / – / –
Tile	99.5 / 97.3 / 90.9	99.8 / 96.3 / 93.6	100.0 / 99.4 / 97.7	99.8 / 98.9 / –	100.0 / 98.7 / –	99.7 / 96.6 / 92.4	– / – / –
Wood	100.0 / 99.2 / 98.1	99.0 / 95.9 / 92.5	99.2 / 98.2 / 90.5	99.7 / 98.9 / –	99.4 / 98.4 / –	99.3 / 95.8 / 93.3	– / – / –
Bottle	100.0 / 98.9 / 96.6	100.0 / 99.0 / 97.1	100.0 / 99.3 / 95.6	100.0 / 98.2 / –	100.0 / 98.9 / –	100.0 / 98.8 / 97.0	– / – / –
Cable	100.0 / 98.5 / 93.7	99.8 / 98.9 / 95.6	99.2 / 98.1 / 93.9	99.8 / 99.3 / –	99.9 / 98.1 / –	99.2 / 98.4 / 93.9	– / – / –
Capsule	100.0 / 99.0 / 94.8	97.7 / 98.4 / 95.4	99.6 / 99.3 / 84.5	98.8 / 97.6 / –	99.5 / 98.5 / –	99.0 / 98.8 / 96.4	– / – / –
Hazelnut	100.0 / 99.0 / 96.8	100.0 / 99.1 / 95.9	100.0 / 99.7 / 93.1	100.0 / 98.9 / –	100.0 / 98.5 / –	100.0 / 99.2 / 96.3	– / – / –
Metal nut	100.0 / 98.7 / 96.5	100.0 / 98.7 / 94.4	99.8 / 98.6 / 94.4	100.0 / 95.8 / –	100.0 / 98.8 / –	100.0 / 98.1 / 93.0	– / – / –
Pill	100.0 / 98.5 / 96.9	98.6 / 99.1 / 97.7	99.1 / 99.0 / 91.0	97.5 / 95.4 / –	98.1 / 97.9 / –	98.4 / 98.3 / 97.0	– / – / –
Screw	100.0 / 99.5 / 97.6	98.0 / 99.6 / 98.6	99.4 / 99.5 / 87.9	98.5 / 98.8 / –	96.9 / 99.1 / –	98.9 / 99.7 / 98.6	– / – / –
Toothbrush	100.0 / 99.1 / 93.4	100.0 / 99.2 / 95.0	100.0 / 98.7 / 91.6	100.0 / 98.9 / –	100.0 / 99.4 / –	100.0 / 99.1 / 94.2	– / – / –
Transistor	100.0 / 97.7 / 94.7	99.7 / 95.4 / 82.3	99.8 / 98.0 / 92.9	100.0 / 99.6 / –	98.3 / 96.2 / –	98.5 / 94.3 / 81.8	– / – / –
Zipper	99.5 / 98.7 / 95.9	99.5 / 98.1 / 94.9	99.6 / 99.2 / 93.4	99.8 / 98.6 / –	98.5 / 97.9 / –	98.6 / 98.8 / 96.3	– / – / –
Mean	99.90 / 98.81 / 96.00	99.46 / 98.41 / 95.20	99.65 / 99.03 / 93.07	99.56 / 98.18 / –	99.30 / 98.62 / 95.31	99.44 / 98.25 / 94.99	99.50 / 98.21 / –

(b) BTAD

Category	**UniNet(Ours)**	ReContrast [10]	RealNet [35]	ReConPatch [14]	PyramidFlow [18]	RD++ [29]	PatchCore [25]
Class 01	100.0 / 97.2 / 81.7	100.0 / 97.0 / 78.6	100.0 / 98.2 / –	99.7 / 96.8 / –	100.0 / 97.4 / –	96.8 / 96.2 / 73.2	98.0 / 96.9 / 64.9
Class 02	93.2 / 96.3 / 60.1	89.5 / 96.2 / 57.0	88.6 / 96.3 / –	87.7 / 96.6 / –	88.2 / 97.6 / –	90.1 / 96.4 / 71.3	81.6 / 95.8 / 47.3
Class 03	100.0 / 99.6 / 98.2	95.7 / 99.3 / 96.5	96.1 / 97.9 / –	100.0 / 99.0 / –	99.3 / 98.1 / –	100.0 / 99.6 / 87.4	99.8 / 99.1 / 67.7
Mean	97.73 / 97.70 / 80.01	95.06 / 97.50 / 77.40	96.07 / 97.90 / –	95.80 / 97.47 / –	95.83 / 97.70 / –	95.63 / 97.41 / 77.30	93.13 / 97.27 / 59.97

(c) MVTec 3D-AD

Category	**UniNet(Ours)**	ReContrast [10]	BTF [13]	Shape-Guided [6]	M3DM [31]	AST [26]	PatchCore [25]
Bagel	100.0 / 95.2	99.1 / –	85.4 / 89.8	91.1 / 94.6	94.4 / 95.2	94.7 / –	91.2 / 89.9
Cable Gland	99.6 / 98.1	95.3 / –	84.0 / 94.8	93.6 / 97.2	91.8 / 97.2	92.8 / –	90.2 / 95.3
Carrot	100.0 / 97.3	92.7 / –	82.4 / 92.7	88.3 / 96.0	89.6 / 97.3	85.1 / –	88.5 / 95.7
Cookie	73.3 / 90.3	69.6 / –	68.7 / 87.2	66.2 / 91.4	74.9 / 89.1	82.5 / –	70.9 / 91.8
Dowel	100.0 / 98.4	97.5 / –	97.4 / 92.7	97.4 / 95.8	95.9 / 93.2	98.1 / –	95.2 / 93.0
Foam	92.8 / 85.4	82.5 / –	71.6 / 55.5	77.2 / 77.6	76.7 / 84.3	95.1 / –	73.3 / 71.9
Peach	98.9 / 98.1	95.0 / –	71.3 / 90.2	78.5 / 93.7	91.9 / 97.0	89.5 / –	72.7 / 92.0
Potato	98.6 / 95.8	67.9 / –	59.3 / 93.1	64.1 / 94.9	64.8 / 95.6	61.3 / –	56.2 / 93.7
Rope	99.9 / 99.0	98.8 / –	92.0 / 90.3	88.4 / 95.6	93.8 / 96.8	99.2 / –	96.2 / 93.8
Tire	94.5 / 97.9	87.9 / –	72.4 / 89.9	70.6 / 95.7	76.7 / 96.6	82.1 / –	76.8 / 92.9
Mean	95.76 / 95.55	88.63 / 95.20	78.52 / 87.63	81.51 / 93.30	85.03 / 94.22	88.00 / –	81.14 / 91.03

(d) The standard deviation

Dataset	△
MVTec AD	0.03
	0.02
	0.04
BTAD	0.01
	0.05
	0.03
MVTec 3D-AD	0.99
	–
	0.33

Table 1. Quantitative results (***one-class***) across three industrial datasets. We report I-AUROC / P-AUROC / PRO on the (a) MVTec AD and (b) BTAD datasets. For the (c) MVTec 3D-AD dataset, P-AUROC is not presented. (d) Each cell, from top to bottom, represents the standard deviation of I-AUROC, P-AUROC, and PRO. Best and second-best results are highlighted in red and blue, respectively.

2.2. Implementation details

UniNet was trained on a computer with NVIDIA GeForce RTX 3090. Following [8, 10], we used the publicly available WideResNet50 pre-trained on ImageNet [9] as S-T models. AdamW [23] was employed as the optimizer with weight decay=1e-5, and the learning rate of 5e-3 and 1e-6 for the learnable student and teacher, respectively. Hyper-parameters n, $\mathcal{T}$, τ, λ, α, and β were set to 3, 2, 1, 0.7, 0.01, and 0.03, respectively. We generally trained for 100 epochs with a batch size of 8, except for three supervised poly datasets (200 epochs), and three unsupervised medical datasets along with the VAD dataset (1000 iterations), saving the best model for evaluation.

All images were resized into 256×256 without data augmentation, except for the MVTec 3D-AD dataset and three polyp segmentation datasets. For the MVTec 3D-AD dataset, only RGB data were used for training and images were first center-cropped before resizing them. Following [5, 30], we adopted a multi-scale $\{0.75, 1, 1.25\}$ training strategy for three polyp segmentation datasets. For a fair comparison, we followed prior works that selected a specific proportion of images from the Kvasir and CVC-ClinicDB datasets for training, while the remaining images were used for testing. For the CVC-ColonDB dataset, we directly employed its training and test sets for training and evaluation.

The procedure for the Weighted Decision Mechanism is outlined in Algorithm 1. The Weighted Decision Mechanism was not applied to the three polyp segmentation datasets, as only segmentation evaluation metrics were con-

(a) MVTec AD

Category	**UniNet(Ours)**	ReContrast [10]	MambaAD [11]	DiAD [12]	DeSTSeg [34]	SimpleNet [22]	UniAD [33]
Carpet	99.4 / 99.8	98.3 / –	99.8 / 99.9	99.4 / 99.9	95.9 / 98.8	95.7 / 98.7	99.8 / 99.9
Grid	99.1 / 99.7	98.9 / –	100.0 / 100.0	98.5 / 99.8	97.9 / 99.2	97.6 / 99.2	98.2 / 99.5
Leather	100.0 / 100.0	100.0 / –	100.0 / 100.0	99.8 / 99.7	99.2 / 99.8	100.0 / 100.0	100.0 / 100.0
Tile	97.8 / 99.2	99.5 / –	98.2 / 99.3	96.8 / 99.9	97.0 / 98.9	99.3 / 99.8	99.3 / 99.8
Wood	100.0 / 100.0	99.7 / –	98.8 / 99.6	99.7 / 100.0	99.9 / 100.0	98.4 / 99.5	98.6 / 99.6
Bottle	100.0 / 100.0	100.0 / –	100.0 / 100.0	99.7 / 96.5	98.7 / 99.6	100.0 / 100.0	99.7 / 100.0
Cable	94.9 / 97.1	95.6 / –	98.8 / 99.2	94.8 / 98.8	89.5 / 94.6	97.5 / 98.5	95.2 / 95.9
Capsule	96.3 / 99.2	97.3 / –	94.4 / 98.7	89.0 / 97.5	82.8 / 95.9	90.7 / 97.9	86.9 / 97.8
Hazelnut	100.0 / 100.0	100.0 / –	100.0 / 100.0	99.5 / 99.7	98.8 / 99.2	99.9 / 99.9	99.8 / 100.0
Metal nut	100.0 / 100.0	100.0 / –	99.9 / 100.0	99.1 / 96.0	92.9 / 98.4	96.9 / 99.3	99.2 / 99.9
Pill	98.3 / 99.6	96.3 / –	97.0 / 99.5	95.7 / 98.5	77.1 / 94.4	88.2 / 97.7	93.7 / 98.7
Screw	100.0 / 100.0	97.2 / –	94.7 / 97.9	90.7 / 99.7	69.9 / 88.4	76.7 / 90.6	87.5 / 96.5
Toothbrush	100.0 / 100.0	96.7 / –	98.3 / 99.3	99.7 / 99.9	71.7 / 89.3	89.7 / 95.7	94.2 / 97.4
Transistor	100.0 / 100.0	94.5 / –	100.0 / 100.0	99.8 / 99.6	78.2 / 79.5	99.2 / 98.7	99.8 / 98.0
Zipper	100.0 / 100.0	99.4 / –	99.3 / 99.8	95.1 / 99.1	88.4 / 96.3	99.0 / 99.7	95.8 / 99.5
Mean	99.05 / 99.64	98.23/ 99.40	98.61 / 99.55	97.15 / 98.97	89.19 / 95.49	95.25 / 98.36	96.51 / 98.83

(b) VisA

Category	**UniNet(Ours)**	ReContrast [10]	MambaAD [11]	DiAD [12]	DeSTSeg [34]	SimpleNet [22]	UniAD [33]
pcb1	100.0 / 100.0	96.5 / –	95.4 / 93.0	88.1 / 88.7	87.6 / 83.1	91.6 / 91.9	92.8 / 92.7
pcb2	99.8 / 99.8	96.8 / –	94.2 / 93.7	91.4 / 91.4	86.5 / 85.8	92.4 / 93.3	87.8 / 87.7
pcb3	92.0 / 93.6	96.8 / –	93.7 / 94.1	86.2 / 87.6	93.7 / 95.1	89.1 / 91.1	78.6 / 78.6
pcb4	100.0 / 100.0	99.9 / –	99.9 / 99.9	99.6 / 99.5	97.8 / 97.8	97.0 / 97.0	98.8 / 98.8
macaroni1	100.0 / 100.0	97.6 / –	91.6 / 89.8	85.7 / 85.2	76.6 / 69.0	85.9 / 82.5	79.9 /79.8
macaroni2	100.0 / 100.0	89.5 / –	81.6 / 78.0	62.5 / 57.4	68.9 / 62.1	68.3 / 54.3	71.6 / 71.6
capsules	99.9 / 100.0	77.7 / –	91.8 / 95.0	58.2 / 69.0	87.1 / 93.0	74.1 / 82.8	55.6 / 55.6
candle	100.0 / 100.0	96.3 / –	96.8 / 96.9	92.8 / 92.0	94.9 / 94.8	84.1 / 73.3	94.1 / 94.0
cashew	96.3 / 98.0	94.5 / –	94.5 / 97.3	91.5 / 95.7	92.0 / 96.1	88.2 / 91.3	92.8 / 92.8
chewinggum	100.0 / 100.0	98.6 / –	97.7 / 98.9	95.1 / 99.5	95.8 / 98.3	96.4 / 98.2	96.3 / 96.2
fryum	99.2 / 99.6	97.3 / –	95.2 / 97.7	89.8 / 95.0	92.1 / 96.1	88.4 / 93.0	83.0 / 83.0
pipe_fryum	99.5 / 99.8	99.3 / –	98.7 / 99.3	96.2 / 98.1	94.1 / 97.1	90.8 / 95.5	94.7 / 94.7
Mean	98.9 / 99.2	95.1/ 96.4	94.3 / 94.5	86.8 / 88.3	88.9 / 89.0	87.2 / 87.0	91.5 / 90.8

Table 2. Quantitative results (***multi-class***) across two industrial datasets. I-AUROC and Image-level AP are reported for the multi-class anomaly detection. Best and second-best results are highlighted in red and blue, respectively.

sidered. For these three polyp datasets, segmentation accuracy was evaluated by comparing the upsampled output of the student model with its pixel-level ground-truth. For the Ped2 dataset, we employed the frame-ped strategy [20], which detects anomalies by measuring the discrepancy between the student-generated frame and its corresponding ground-truth.

3. More experimental results

3.1. Results on the industrial datasets

Traditional methods develop separate models for each category, known as the *one-class* anomaly detection setting. Recent efforts [10, 11, 33] have attempted to design a unified model that can handle multiple categories, *i.e.*, the *multi-class* anomaly detection setting. Experimental results for both settings are reported as follows.

Results under the one-class setting. In addition to the overall average results across all categories from the MVTec AD, BTAD, and MVTec 3D-AD datasets, the average results for each individual category from these three datasets are also presented in Table 1. As reported in Table 1(a), UniNet achieves **100.0**% anomaly detection performance across all categories of the MVTec AD dataset, with the exception for the grid, tile, and zipper categories. It also shows comparable segmentation performance. Moreover, UniNet demonstrates notable anomaly detection and segmentation performance across most categories in the other two datasets, as illustrated in Table 1(b) and (c). Particularly, on the MVTec 3D-AD dataset, UniNet achieves the best and significant results across all categories, except for *Cookie* category. Finally, the standard deviations of the three evaluation metrics across three datasets are presented in Table 1(d).

Results under the multi-class setting. Table 2 shows the multi-class anomaly detection on the MVTec AD and VisA [36] datasets. Following [10–12], both I-AUROC and Image-level AP are reported. UniNet was compared with state-of-the-art methods reported in [11]: MambaAD [11], DiAD [12], DeSTSeg [34], SimpleNet [22], and UniAD [33]. As shown in Table 2(a), UniNet similarly shows strong performance across most categories, achieving the

highest average I-AUROC and Image-level AP, with a perfect score of **100.0**%. UniNet outperforms the other methods and the baseline model by 0.44% and 0.82% in I-AUROC, as well as by 0.09% and 0.24% in Image-level AP. Moreover, as reported in Table 2(b), UniNet also achieves impressive anomaly detection performance on the more challenging VisA dataset, obtaining the best results in every category except for the *pcb3* category. UniNet markedly surpasses leading methods, with improvement of **3.8%** in I-AUROC and **2.8%** in Image-level AP, respectively.

3.2. Complexity analysis

Table 3 investigates the complexity of UniNet and the baseline model, ReContrast [10]. By utilizing the same backbone as the baseline model, UniNet achieves a comparable model size to ReContrast, while offering a higher inference speed with an improvement of 6.77 FPS. Additionally, UniNet outperforms ReContrast in terms of I-AUROC, P-AUROC and PRO, with increases of 0.44%, 0.40%, and 0.80%, respectively.

Method	Model Size (GB) ↓	Speed (FPS) ↑	Infer. Time (s) ↓	Metrics
ReContrast	**0.141**	9.46	15.6	99.46 / 98.41 / 95.20
UniNet	0.150	**16.23**	**8.2**	**99.90 / 98.81 / 96.00**

Table 3. Complexity analysis between UniNet and the baseline model on the MVTec AD dataset. Metrics are I-AUROC / P-AUROC / PRO. Best results are highlighted in bold.

4. Supplementary ablation studies

4.1. Study on Multi-Scale Embedding Module

To validate the effectiveness of MEM within the bottleneck, we studied the effect on the kernel size k in MEM. The results are presented in Table 4. Both detection and segmentation performance steadily improve as the large kernel size increases and the best results can be obtained when using a combination of (3, 7). Notably, larger kernels lead to a higher number of model size and decreased inference speed. As shown in Table 4, the model size of the bottleneck (*e.g.*, 0.179 GB and 0.363 GB) can significantly surpass that of the entire framework (see Table 3) prior to re-parameterization. Similarly, as the kernel size increases, the inference speed of UniNet progressively decreases. However, re-parameterization results in a smaller model size and improved inference speed.

4.2. Study on Domain-Related Feature Selection

To demonstrate that introducing domain-related information into the student aids in improving its feature representations, we investigated the impact of different selection strategies on three datasets, as shown in Table 5. Without selecting representative features from the teacher, the student faces challenges in understanding target-oriented feature information, especially on more structurally complex datasets (*e.g.*, VAD), which negatively affects performance. Conversely, our method effectively guides the student to select and learn the most crucial features, yielding promising results. However, selecting all available information from the teacher may not be beneficial, as it could include unimportant details.

We also investigated the impacts of DFS on the performance of ReContrast, as illustrated in Fig. 1. Without DFS, ReContrast fails to sufficiently learn vital domain-related features, leading to the loss of subtle details–such as the label on a pill being mistakenly identified as an anomaly. By incorporating DFS, ReContrast mitigate this issue by selecting key features for learning.

k	Model Size (GB) ↓	Speed (FPS) ↑	Metrics
(3, 3)	**0.077**+0.00%	**16.04**+0.00%	99.82 / 98.16 / 95.81
(3, 5)	0.118-34.75%	15.86+2.28%	99.88 / 98.20 / 95.87
(3, 7)	0.179-57.02%	15.33+5.55%	**99.90 / 98.81 / 96.00**
(3, 11)	0.363-78.80%	12.13+25.27%	99.87 / 98.16 / 95.75

Table 4. Study on kernel size k in MEM on MVTec AD dataset, with only the model size of bottleneck reported. Metrics are I-AUROC / P-AUROC / PRO. The gains after re-parameterization are highlighted in green, with the best results indicated in bold.

Method	Dataset		
	MVTec AD	APTOS	VAD
F_S	99.77 / 98.76 / 95.73	99.99 / **99.63** / **99.50**	99.25 / 95.88 / 95.90
$(F_S)^P$	**99.90 / 98.81 / 96.00**	**100.0** / 99.60 / 99.44	**99.95 / 98.60 / 98.60**
$(F_S)^A$	99.81 / 98.75 / 95.80	99.99 / 99.55 / 99.37	99.87 / 98.20 / 98.20

Table 5. Study on different selection strategies for DFS. For the MVTec AD, the evaluation metrics include I-AUROC / P-AUROC / PRO. For the APTOS and VAD datasets, three metrics are reported: I-AUROC / FI / ACC. "F_S", "$(F_S)^P$", and "$(F_S)^A$" refer to no feature selection, selecting representative features, and selecting all available features, respectively. Best results are highlighted in bold.

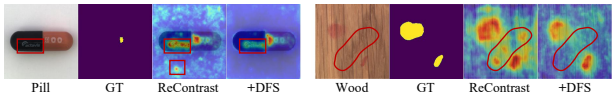

Figure 1. Segmentation results w/o and w DFS on the MVTec AD dataset.

4.3. Additional study on key elements

In addition to industrial datasets, ablation studies on the key components of UniNet on medical and video domains are listed in Table 6.

MEM	DFS	$\mathcal{L}_{SC}$	$\mathcal{L}_M$	$\mathcal{M}$	ATPOS	Ped2
					95.17	95.01
✓					95.79	95.30
✓	✓				96.52	95.63
✓		✓			96.24	95.35
✓	✓	✓			97.89	96.09
	✓	✓	✓		97.80	96.20
✓		✓	✓	✓	99.55	97.40
✓	✓	✓	✓	✓	**100.0**	**97.91**

Table 6. Ablation studies on the key elements of UniNet on medical and video datasets, with I-AUROC listed.

$\mathcal{T}$	α	β	ATPOS	Ped2
0.1	0.01	0.03	99.88	97.72
0.1	0.1	0.1	99.85	97.66
0.5	0.05	0.05	99.90	**97.94**
0.5	0.01	0.05	99.90	97.84
1	0.1	0.03	99.60	97.68
1	0.01	0.1	99.72	97.60
2	0.01	0.03	**100.0**	97.91
2	0.1	0.05	99.97	97.80

Table 7. Hyper-parameter analysis on medical and video datasets, with I-AUROC reported.

4.4. Hyper-parameter sensitivity analysis

The main hyper-parameters include temperature coefficient $\mathcal{T}$ and $\{\alpha, \beta\}$ (controlling the upper and lower limits of the weight in $\mathcal{M}$). As shown in Table 7, we evaluated different combinations of $\mathcal{T}$ (0.1, 0.5, 1, 2), α (0.01, 0.05, 0.1), and β (0.03, 0.05, 0.1).

5. Qualitative Results

To clearly validate the superior segmentation performance of UniNet, comprehensive visualization results are presented across three industrial datasets and three medical datasets.

5.1. Visualization on industrial datasets

As illustrated in Fig. 2, UniNet effectively segments both local and global anomalies across texture and object categories, while maintaining lower anomaly scores in regions devoid of anomalies.

Results on the BTAD and MTVec 3D-AD datasets are respectively shown in Fig. 3(a) and (b). For the BTAD dataset, despite the anomalies closely resembling normal areas, UniNet exhibits exceptional segmentation performance, effectively detecting even the smallest anomalies. For the MVTec 3D-AD dataset, using only the RGB modality, UniNet still achieves promising segmentation results, as shown in Fig. 3(b). However, due to lack of multi-modal information, UniNet may fail to maintain lower anomaly scores in some normal regions, such as the *Bagel*, *Cookie*, and *Potato* categories. This is because the chocolates in the *Cookie* category resemble anomalies, such as holes. As a result, relying on a single modality alone makes it challenge to achieve more accurate segmentation. We will explore combining other modalities with the RGB modality later.

5.2. Visualization on medical datasets

In addition to industrial datasets, results on three polyp datasets are visualized in Fig. 3(c). Despite the variability in images collected from the intestinal environments of different patients, UniNet also demonstrates superior segmentation performance in polyps. As illustrated in Fig 3(c), the segmented results perfectly match the ground-truths, demonstrating that UniNet is highly resistant to both over-segmentation and under-segmentation.

6. Discussion

6.1. Limitation

Similar to ReContrast [10] and other unsupervised AD methods [8, 17, 22], UniNet also experiences training instability for certain categories, with performance fluctuating when overtraining occurs or random seeds are changed, particularly in anomaly segmentation performance. However, thank to weighted decision mechanism $\mathcal{M}$, anomaly detection performance can hardly be influenced, ensuring robust anomaly detection results. Besides, although UniNet has achieved promising results on multimodal datasets like MVTec 3D-AD, relying solely on 2D data limits its potential for better anomaly detection performance.

6.2. Future work

We will apply UniNet to other tasks, such as multimodal anomaly detection or 3D medical image segmentation, by incorporating other modalities like text or point cloud to achieve superior performance. Also, the optimization of loss functions and the model will be investigated to ensure more stable training.

References

[1] Aimira Baitieva, David Hurych, Victor Besnier, and Olivier Bernard. Supervised anomaly detection for complex industrial images. In *CVPR*, pages 17754–17762, 2024. 1

[2] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *CVPR*, pages 9592–9600, 2019. 1

[3] Paul Bergmann, Xin Jin, David Sattlegger, and Carsten Steger. The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. *arXiv preprint arXiv:2112.09045*, 2021. 1

[4] Jorge Bernal, F Javier Sánchez, Gloria Fernández-Esparrach, Debora Gil, Cristina Rodríguez, and Fernando Vilariño. Wm-dova maps for accurate polyp highlighting

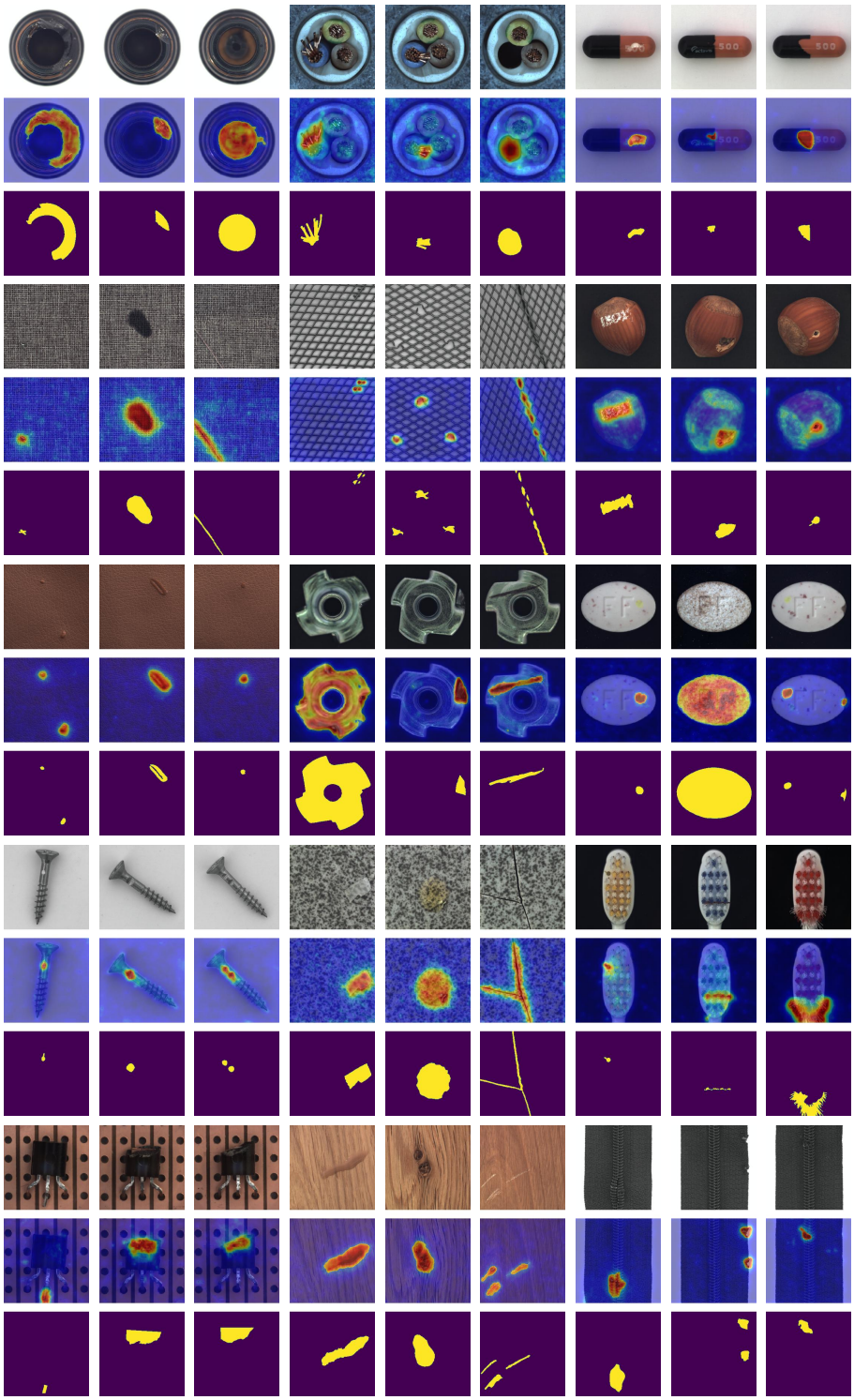

Figure 2. Visualization of UniNet on the MVTec AD dataset. Each group, from top to bottom, displays the anomalous images, our segmentation results, and ground-truths, respectively.

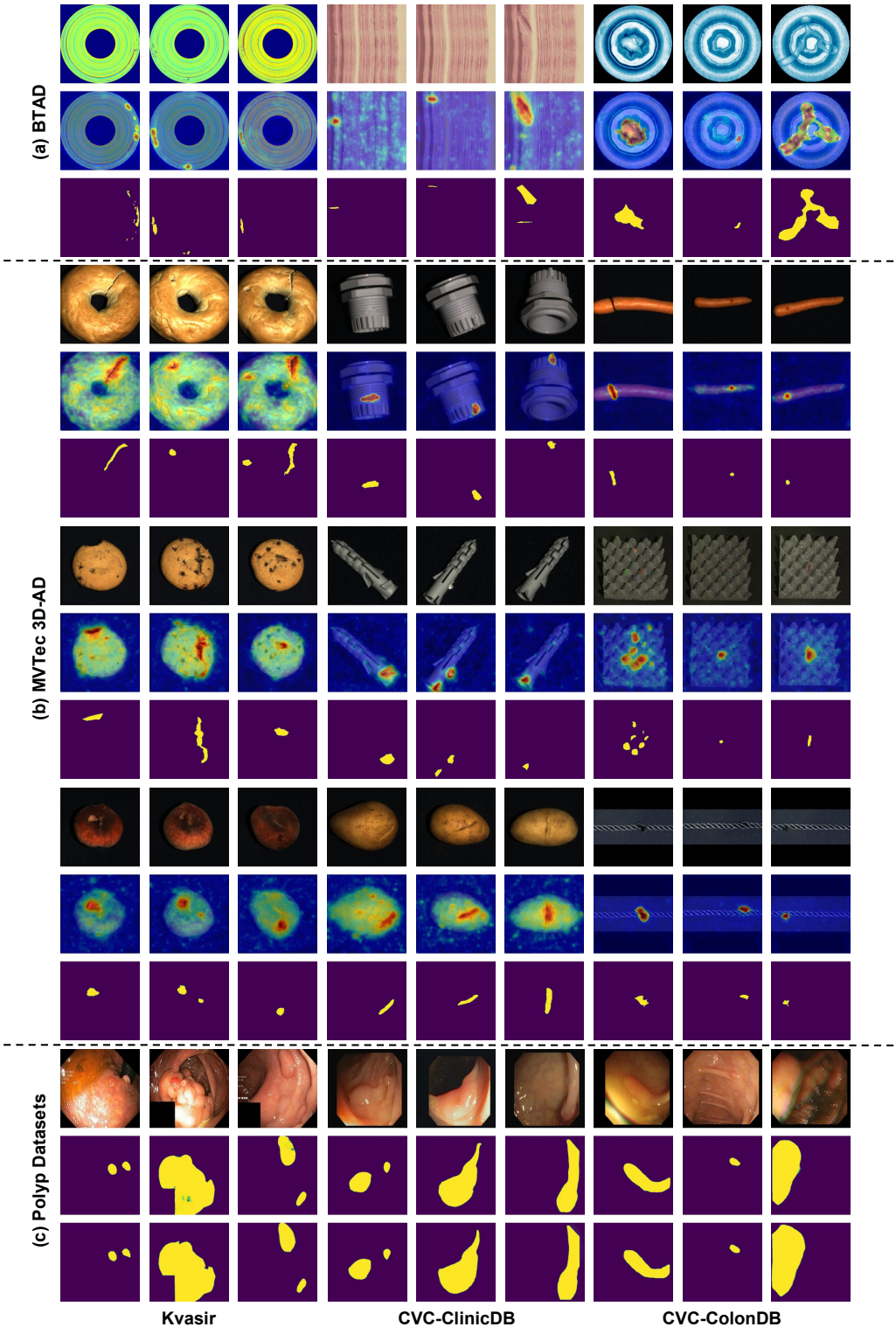

Figure 3. Visualization of UniNet on the BTAD, MVTec 3D-AD, and three polyp datasets (Kvasir, CVC-ClinicDB, and CVC-ColonDB). For the MVTec 3D-AD dataset, all categories are listed in order: *Bagel, Cable gland, Carrot, Cookie, Dowel, Foam, Peach, Potato,* and *Rope*. Each group, from top to bottom, displays the anomalous images, our segmentation results, and ground-truths, respectively.

in colonoscopy: Validation vs. saliency maps from physicians. *Computerized medical imaging and graphics*, 43:99–111, 2015. 1

[5] Nhat-Tan Bui, Dinh-Hieu Hoang, Quang-Thuc Nguyen, Minh-Triet Tran, and Ngan Le. Meganet: Multi-scale edge-guided attention network for weak boundary polyp segmentation. In *WACV*, pages 7985–7994, 2024. 2

[6] Yu-Min Chu, Chieh Liu, Ting-I Hsieh, Hwann-Tzong Chen, and Tyng-Luh Liu. Shape-guided dual-memory learning for 3d anomaly detection. In *ICML*, pages 6185–6194, 2023. 2

[7] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, et al. Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic). *arXiv preprint arXiv:1902.03368*, 2019. 1

[8] Hanqiu Deng and Xingyu Li. Anomaly detection via reverse distillation from one-class embedding. In *CVPR*, pages 9737–9746, 2022. 2, 5

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, pages 248–255, 2009. 2

[10] Jia Guo, Shuai Lu, Lize Jia, Weihang Zhang, and Huiqi Li. Recontrast: Domain-specific anomaly detection via contrastive reconstruction. In *NIPS*, pages 10721–10740, 2023. 1, 2, 3, 4, 5

[11] Haoyang He, Yuhu Bai, Jiangning Zhang, Qingdong He, Hongxu Chen, Zhenye Gan, Chengjie Wang, Xiangtai Li, Guanzhong Tian, and Lei Xie. Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. *arXiv preprint arXiv:2404.06564*, 2024. 3

[12] Haoyang He, Jiangning Zhang, Hongxu Chen, Xuhai Chen, Zhishan Li, Xu Chen, Yabiao Wang, Chengjie Wang, and Lei Xie. A diffusion-based framework for multi-class anomaly detection. In *AAAI*, pages 8472–8480, 2024. 3

[13] Eliahu Horwitz and Yedid Hoshen. Back to the feature: classical 3d features are (almost) all you need for 3d anomaly detection. In *CVPR*, pages 2968–2977, 2023. 2

[14] Jeeho Hyun, Sangyun Kim, Giyoung Jeon, Seung Hwan Kim, Kyunghoon Bae, and Byung Jun Kang. Reconpatch: Contrastive patch representation learning for industrial anomaly detection. In *WACV*, pages 2052–2061, 2024. 2

[15] Debesh Jha, Pia H Smedsrud, Michael A Riegler, Pål Halvorsen, Thomas De Lange, Dag Johansen, and Håvard D Johansen. Kvasir-seg: A segmented polyp dataset. In *MultiMedia modeling: 26th international conference, MMM 2020, Daejeon, South Korea, January 5–8, 2020, proceedings, part II 26*, pages 451–462, 2020. 1

[16] Daniel S Kermany, Michael Goldbaum, Wenjia Cai, Carolina CS Valentim, Huiying Liang, Sally L Baxter, Alex McKeown, Ge Yang, Xiaokang Wu, Fangbing Yan, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *cell*, 172(5):1122–1131, 2018. 1

[17] Sungwook Lee, Seunghyun Lee, and Byung Cheol Song. Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access*, 10: 78446–78454, 2022. 5

[18] Jiarui Lei, Xiaobo Hu, Yue Wang, and Dong Liu. Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In *CVPR*, pages 14143–14152, 2023. 2

[19] Weixin Li, Vijay Mahadevan, and Nuno Vasconcelos. Anomaly detection and localization in crowded scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(1):18–32, 2013. 1

[20] Wen Liu, Weixin Luo, Dongze Lian, and Shenghua Gao. Future frame prediction for anomaly detection–a new baseline. In *CVPR*, pages 6536–6545, 2018. 3

[21] Wenrui Liu, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Diversity-measurable anomaly detection. In *CVPR*, pages 12147–12156, 2023. 2

[22] Zhikang Liu, Yiming Zhou, Yuansheng Xu, and Zilei Wang. Simplenet: A simple network for image anomaly detection and localization. In *CVPR*, pages 20402–20411, 2023. 3, 5

[23] Ilya Loshchilov, Frank Hutter, et al. Fixing weight decay regularization in adam. *arXiv preprint arXiv:1711.05101*, 5, 2017. 2

[24] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In *ISIE*, pages 01–06, 2021. 1

[25] Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *CVPR*, pages 14318–14328, 2022. 2

[26] Marco Rudolph, Tom Wehrbein, Bodo Rosenhahn, and Bastian Wandt. Asymmetric student-teacher networks for industrial anomaly detection. In *WACV*, pages 2592–2602, 2023. 2

[27] Asia Pacific Tele-Ophthalmology Society. Aptos 2019 blindness detection. 2019. 1

[28] Nima Tajbakhsh, Suryakanth R Gurudu, and Jianming Liang. Automated polyp detection in colonoscopy videos using shape and context information. *IEEE Trans. medical imaging*, 35(2):630–644, 2015. 1

[29] Tran Dinh Tien, Anh Tuan Nguyen, Nguyen Hoang Tran, Ta Duc Huy, Soan Duong, Chanh D Tr Nguyen, and Steven QH Truong. Revisiting reverse distillation for anomaly detection. In *CVPR*, pages 24511–24520, 2023. 2

[30] Quoc-Huy Trinh, Minh-Van Nguyen, and Phuoc-Thao Vo Thi. Kdas: Knowledge distillation via attention supervision framework for polyp segmentation. In *ICME*, pages 1–6, 2024. 2

[31] Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, and Chengjie Wang. Multimodal industrial anomaly detection via hybrid fusion. In *CVPR*, pages 8032–8041, 2023. 2

[32] Hang Yao, Ming Liu, Haolin Wang, Zhicun Yin, Zifei Yan, Xiaopeng Hong, and Wangmeng Zuo. Glad: Towards better reconstruction with global and local adaptive diffusion models for unsupervised anomaly detection. *arXiv preprint arXiv:2406.07487*, 2024. 2

[33] Zhiyuan You, Lei Cui, Yujun Shen, Kai Yang, Xin Lu, Yu Zheng, and Xinyi Le. A unified model for multi-class anomaly detection. *NIPS*, 35:4571–4584, 2022. 3

[34] Xuan Zhang, Shiyu Li, Xi Li, Ping Huang, Jiulong Shan, and Ting Chen. Destseg: Segmentation guided denoising student-teacher for anomaly detection. In *CVPR*, pages 3914–3923, 2023. 3

[35] Ximiao Zhang, Min Xu, and Xiuzhuang Zhou. Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In *CVPR*, pages 16699–16708, 2024. 2

[36] Yang Zou, Jongheon Jeong, Latha Pemula, Dongqing Zhang, and Onkar Dabeer. Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In *ECCV*, pages 392–408. Springer, 2022. 1, 3