Learning to Detect Multi-class Anomalies with Just One Normal Image Prompt

Bin-Bin Gao, Tencent YouTu Lab, csgaobb@gmail.com

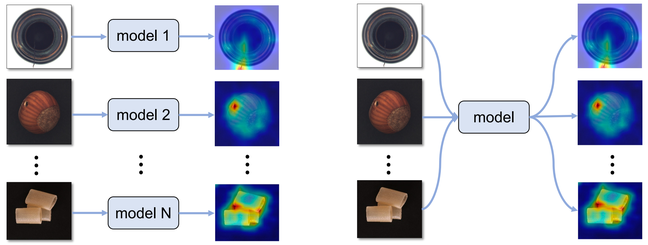
Tencent YouTu Lab

OneNIP Code

ECCV MILANO 2024

1. Background

Unified Multi-Class Anomaly Detection

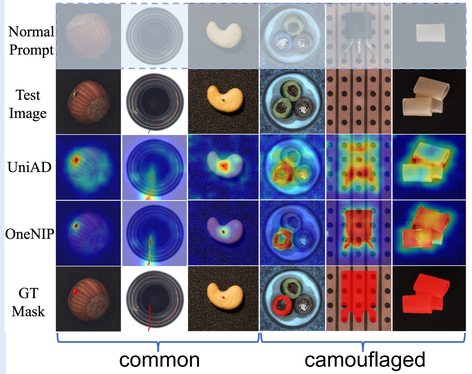


Separated Paradigm
one model for one class

Unified Paradigm
one model for all classes

The unified AD paradigm attempts to detect multi-class anomalies using a single model. Compared to the separated mode, the unified AD paradigm is more challenging as it requires handling more complex data distributions.

2. Motivation



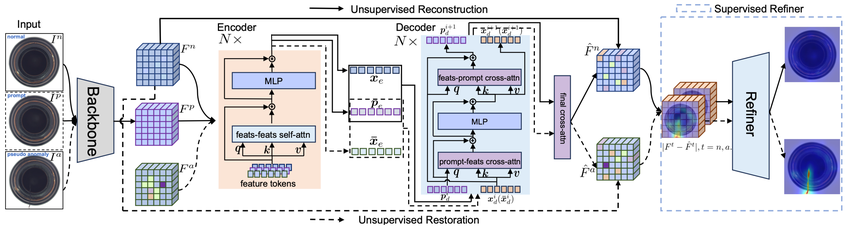
The common anomalies can be detected using their own contextual information.

The camouflaged anomalies are, however, hard to detect if the corresponding normal image prompts are absent.

Inspiration: how to effectively utilize normal image prompts to facilitate anomaly detection in unified model?

3. Our Method: OneNIP

The Pipeline of Our OneNIP Framework



OneNIP is built on the state-of-the-art UniAD, which mainly consists of unsupervised reconstruction, unsupervised restoration, and supervised refiner.

3.1 Reconstruction with Normal Image Prompt

learnable query embedding

$$\boldsymbol{q}' = \mathrm{softmax}(\boldsymbol{q}^i \boldsymbol{x}_e^T/\sqrt{c})\boldsymbol{x}_e,$$
$$\boldsymbol{x}_d^{i+1} = \mathrm{softmax}(\boldsymbol{q}' \boldsymbol{x}_d^{i\,T}/\sqrt{c})\boldsymbol{x}_d^i,$$
$$\boldsymbol{x}_d^0 = \boldsymbol{x}_e, \quad \text{query tokens}$$

Eq.1 LQD in UniAD

normal prompt token

$$\boldsymbol{q}' = \mathrm{softmax}(\boldsymbol{p}_e \boldsymbol{x}_e^T/\sqrt{c})\boldsymbol{x}_e,$$
$$\boldsymbol{x}_d^{i+1} = \mathrm{softmax}(\boldsymbol{q}' \boldsymbol{x}_d^{i\,T}/\sqrt{c})\boldsymbol{x}_d^i,$$
$$\boldsymbol{x}_d^0 = \boldsymbol{x}_e, \quad \text{query tokens}$$

Eq.2 unidirectional decoder with static prompt

$$\text{prompt-to-feature:} \quad \boldsymbol{p}_d^{i+1} = \mathrm{softmax}(\boldsymbol{p}_d^i \boldsymbol{x}_d^{i\,T}/\sqrt{c})\boldsymbol{x}_d^i,$$
$$\text{feature-to-prompt:} \quad \boldsymbol{x}_d^{i+1} = \mathrm{softmax}(\boldsymbol{x}_d^i \boldsymbol{p}_d^{i+1\,T}/\sqrt{c})\boldsymbol{p}_d^{i+1},$$
$$\text{normal prompt token} \quad \boldsymbol{p}_d^0 = \boldsymbol{p}_e, \ \boldsymbol{x}_d^0 = \boldsymbol{x}_e, \quad \text{query tokens}$$

Eq.3 bidirectional decoder with dynamic prompt

3.2 Restoration with Normal Image Prompt

The bidirectional decoder is initialized by $\boldsymbol{p}_e$ and $\bar{\boldsymbol{x}}_e$ (pseudo anomaly token), and dynamically updated with Eq. 3.

3.3 Supervised Refiner

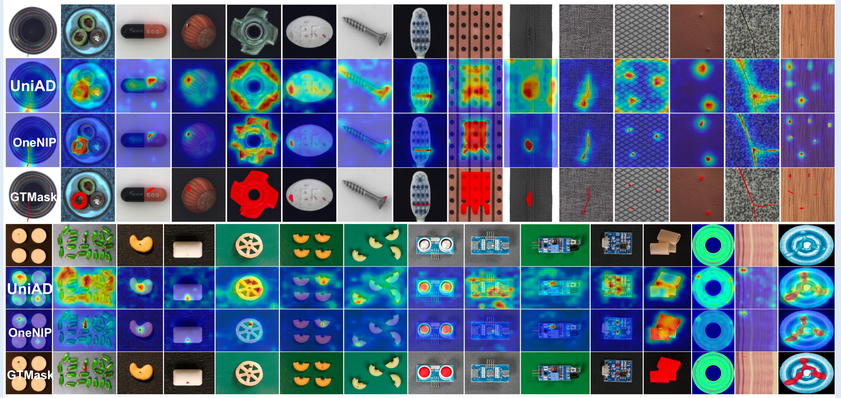
We design a lightweight and pixel-level refiner based on reconstruction errors of normal and pseudo anomaly tokens for achieving more accurate anomaly segmentation.

4. Experiments

4.1 Comparisons with State-of-the-Arts

Datasets	Metric↑	Embedding-based					Discriminator-based		Reconstruction-based	
		CS-Flow [38]	PaDiM [10]	DFM [1]	PatchCore [37]	CFA [20]	DRAEM [54]	SimpleNet [25]	UniAD [52]	OneNIP
MVTec [4]	I-ROC/PR	81.4 / 90.2	87.5 / 92.8	69.7 / 89.8	89.8 / 96.3	80.4 / 91.0	91.4 / 95.3	78.2 / 90.0	96.5 / 98.9	97.9 / 99.3
	P-ROC/PR	93.8 / 33.8	95.5 / 37.8	96.5 / 42.4	96.4 / 50.1	90.7 / 37.1	85.2 / 49.6	81.0 / 24.8	96.8 / 44.7	97.9 / 63.7
BTAD [26]	I-ROC/PR	91.8 / 96.3	95.7 / 97.4	68.8 / 82.8	89.2 / 96.4	87.5 / 87.7	84.7 / 95.0	90.3 / 95.0	92.2 / 97.9	92.6 / 98.5
	P-ROC/PR	95.9 / 34.6	96.7 / 48.7	96.3 / 48.0	96.3 / 48.4	95.6 / 40.4	74.2 / 12.3	78.8 / 36.2	97.1 / 50.9	97.4 / 56.8
VisA [60]	I-ROC/PR	75.8 / 80.0	78.1 / 78.3	51.6 / 77.8	90.3 / 92.0	69.0 / 73.8	81.8 / 85.8	89.2 / 92.2	90.8 / 93.0	92.5 / 94.5
	P-ROC/PR	95.6 / 18.6	95.9 / 17.1	96.5 / 25.2	96.8 / 38.2	91.4 / 16.8	78.1 / 15.1	95.3 / 33.1	98.4 / 33.6	98.7 / 43.3

4.2 Qualitative Comparisons



4.3 Ablation Studies

Table 4: Ablation studies on MVTec. Default settings are marked in blue.

(a) Prompt strategy in Reconstruction, Restoration, and Refiner

No.	Prompt	Res.	Ref.	I-ROC	P-ROC	I-PR	P-PR
0	✗	✗	✗	96.5	96.8	98.9	44.7
1	static	✗	✗	96.8	97.0	98.9	45.8
2	dynamic	✗	✗	97.5	97.1	99.2	46.0
3	✗	✓	✗	96.7	97.0	98.9	46.5
4	dynamic	✓	✗	97.4	97.3	99.1	48.4
5	dynamic	✓	✓	**97.9**	**97.9**	**99.3**	**63.7**

(b) Effects of the number of Encoder, and Decoder

Enc	Dec	I-ROC	P-ROC	I-PR	P-PR
1	1	94.8	97.0	97.9	56.0
2	2	96.7	97.4	98.9	59.4
4	4	97.9	97.9	99.3	63.7
6	6	**98.1**	**98.0**	**99.4**	**64.6**
2	4	97.0	97.6	99.0	61.2
4	2	97.1	97.6	99.0	62.1

(c) Effects of weight α

α	I-ROC	P-ROC	I-PR	P-PR
0.00	97.6	97.3	99.2	48.3
0.25	97.8	97.7	99.3	59.3
0.50	97.9	97.9	99.3	63.7
1.00	96.7	96.7	98.9	63.7

(d) Different prompt modes of the same category

Train	Test	I-ROC	P-ROC	I-PR	P-PR
rand	rand	97.85±0.01	97.86±0.00	99.27±0.01	63.71±0.01
	fixed	97.85±0.02	97.86±0.00	99.27±0.01	63.71±0.02
fixed	fixed	97.91	97.86	99.30	63.66
	rand	96.05±0.24	97.49±0.03	98.34±0.19	60.65±0.18

performance on a more complex data distribution

Datasets	#Classes	Metric ↑	UniAD [52]	OneNIP
MVTec [4]	15	I-ROC/PR	94.8/98.0	97.1/99.0
		P-ROC/PR	96.2/42.1	97.6/61.1
BTAD [26]	3	I-ROC/PR	92.0/97.1	92.0/97.5
		P-ROC/PR	97.1/48.0	97.9/59.0
VisA [60]	12	I-ROC/PR	89.9/92.4	91.9/93.9
		P-ROC/PR	98.3/33.2	98.6/40.6
All	30	I-ROC/PR	92.6/95.7	94.5/96.8
		P-ROC/PR	97.1/39.1	98.0/52.4

comparison results of OneNIP with different resolution

Datasets	Metric ↑	224×224	256×256	320×320
MVTec [4]	I-ROC/PR	97.9/99.3	97.6/99.2	97.9/99.3
	P-ROC/PR	97.9/63.7	97.8/64.7	97.9/65.9
BTAD [26]	I-ROC/PR	92.6/98.5	94.9/99.0	95.3/98.9
	P-ROC/PR	97.4/56.8	97.6/57.0	97.8/57.6
VisA [60]	I-ROC/PR	92.5/94.5	93.3/94.3	94.2/95.7
	P-ROC/PR	98.7/43.3	98.8/44.1	98.8/46.1