Beyond Normal References: Discriminative Few-Shot Anomaly Detection

Huan Wang[1,2] Jun Shen[1] Jun Yan[1] Guansong Pang[2]
1 University of Wollongong, Australia 2 Singapore Management University, Singapore

IDEAL · NeurIPS 2026

Code

1 Problem & fixed-reference protocol

Few references, fixed across all target queries.

Sparse anomaly coverage challenges unseen detection.



General
Normal + seen + unseen

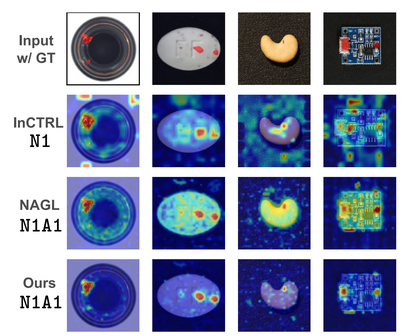
Hard
Normal + unseen types only

Why direct similarity matching fails

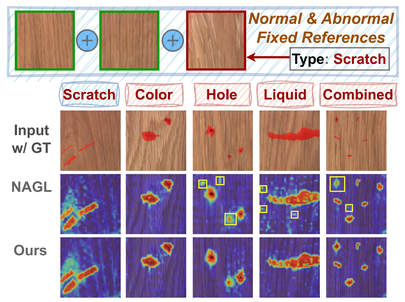


Different reference/query defects → noisy matches. (Fig. 1)

Qualitative localization & unseen defects

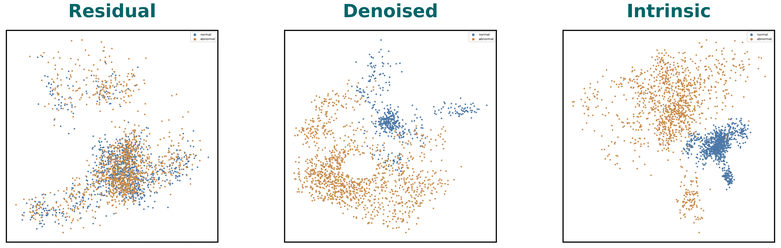


Diverse industrial objects



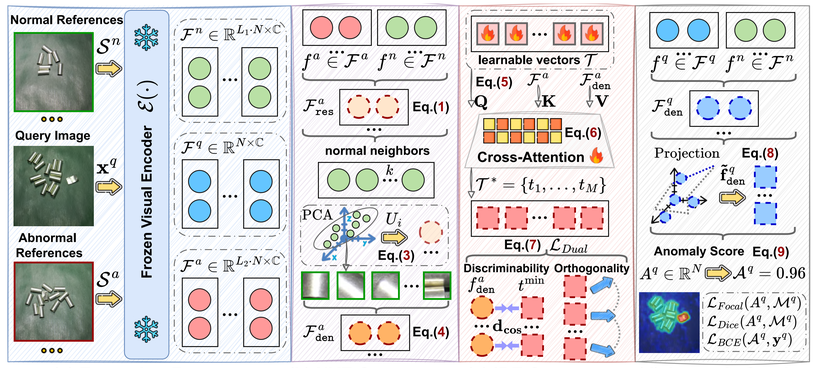
Fixed scratch reference · N2A1

Deviation representations (t-SNE)



VisA, trained on MVTecAD. Normal / abnormal patches. (Fig. 3)

2 Method: intrinsic deviation learning



Paper framework (Eqs. 1–9): features → NVE → IDE → scoring.

Normal Variation Eraser (NVE)

$$\mathcal{F}^a_{\text{res}} = \{f^a_i - \texttt{argmin}_{f^n \in \mathcal{F}^n} \mathbf{d}_{\cos}(f^n, f^a_i)\}_{i=1}^{L_2 \cdot N} \in \mathbb{R}^{L_2 \cdot N \times \mathbb{C}}. \quad (1)$$

$$\mathcal{F}^a_{\text{den}} = \{f^a_{i,\text{res}} - \alpha U_i U_i^\top f^a_{i,\text{res}}\}_{i=1}^{L_2 \cdot N} = \{f^a_{i,\text{den}}\}_{i=1}^{L_2 \cdot N} \in \mathbb{R}^{L_2 \cdot N \times \mathbb{C}}, \quad (4)$$

Local PCA: k = 12 neighbors, rank 4, α = 0.8.

Intrinsic Deviation Encoder (IDE)

$$\mathbf{Q} = \omega^Q(\mathcal{T}) \in \mathbb{R}^{M \times \mathbb{C}},\ \mathbf{K} = \omega^K(\mathcal{F}^a) \in \mathbb{R}^{L_2 \cdot N \times \mathbb{C}},\ \mathbf{V} = \omega^V(\mathcal{F}^a_{\text{den}}) \in \mathbb{R}^{L_2 \cdot N \times \mathbb{C}}, \quad (5)$$

$$\mathcal{T}^* = \{t_m\}_{1 \le m \le M} = \texttt{FFN}(\texttt{Softmax}(\mathbf{Q}\mathbf{K}^\top / \sqrt{\mathbb{C}} + \gamma(1 - \mathcal{M}^a))\mathbf{V}), \quad (6)$$

$$\mathcal{L}_{Dual} = \lambda_1 \underbrace{\frac{1}{\sum_{i=1}^{L_2 \cdot N} \mathcal{M}^a_i} \sum_{i=1}^{L_2 \cdot N} \mathcal{M}^a_i \mathbf{d}_{\cos}(f^a_{i,\text{den}}, t^{\min}_i)}_{\text{Discriminability}} + \lambda_2 \underbrace{\frac{1}{M(M-1)} \sum_{m_1 \neq m_2} \left(\frac{t_{m_1} \cdot t_{m_2}}{\|t_{m_1}\| \times \|t_{m_2}\|}\right)^2}_{\text{Orthogonality}}, \quad (7)$$

45 vectors · γ → −∞ masks normal reference patches.

Projection-based anomaly scoring

$$\tilde{\mathbf{f}}^q_{i,\text{den}} = \texttt{Proj}(f^q_{i,\text{den}}) := \sum_{m=1}^{M} (\langle f^q_{i,\text{den}}, t_m \rangle / \langle t_m, t_m \rangle) t_m, \quad (8)$$

$$A^q_i = \frac{1}{2}(1 - \mathbf{d}_{\cos}(f^q_{i,\text{den}}, \tilde{\mathbf{f}}^q_{i,\text{den}}) + \mathbf{d}_{\cos}(f^q_i, f^n_{i,\min})) \ \Rightarrow\ A^q = \{A^q_i\}_{i=1}^{N} \in \mathbb{R}^N, \quad (9)$$

Image score = mean of top 1% patch scores.

d cos: cosine distance.

Training & deployment

$$\mathcal{L}_{train} = \mathcal{L}_{Focal}(A^q, \mathcal{M}^q) + \mathcal{L}_{Dice}(A^q, \mathcal{M}^q) + \mathcal{L}_{BCE}(\mathcal{A}^q, \mathbf{y}^q) + \mathcal{L}_{Dual}. \quad (10)$$

Frozen encoder; only IDE is trained on auxiliary episodes.
Test time: no tuning; one abnormal-reference mask reused.

3 Experiments

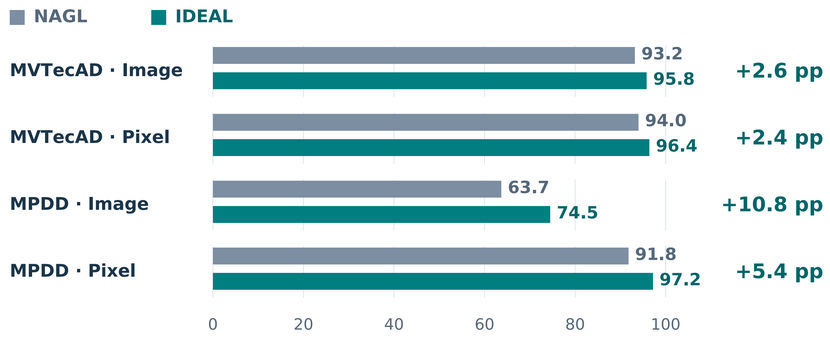
General · N1A1 · image / pixel AUROC (%)

Dataset	NAGL	IDEAL
MVTecAD	95.1 / 96.1	**96.3 / 96.8**
VisA	88.5 / 97.5	**90.8 / 97.8**
AITEX	71.6 / 75.5	**75.8 / 82.7**
MPDD	77.1 / 96.5	**78.5 / 97.9**
BTAD	92.0 / 95.8	**93.4 / 97.0**
BraTS	67.7 / 93.7	**74.9 / 94.8**
Liver	59.9 / 94.8	**62.7 / 97.2**
RESC	78.9 / 91.7	**82.8 / 92.5**

VisA → MVTecAD; MVTecAD → other datasets.
Mean of 3 runs. No target tuning. (Tables 1, A9–A10)

Unseen anomaly types: Hard setting

N1A1 · reference anomaly type excluded from test queries



Component ablation

MVTecAD · N1A1 · General · AUROC (%)

Configuration	Image	Pixel
Both-reference matching	90.1	92.2
NVE only (no IDE)	93.9	94.3
IDE only (no NVE)	94.5	96.0
NVE + IDE (full)	**96.3**	**96.8**

21.9 FPS	**0.5 h**	**23.83 M**
inference	training	parameters

Efficiency benchmark: VisA. (Tables 2–3)