ResAD: A Simple Framework for Class Generalizable Anomaly Detection

NeurIPS 2024

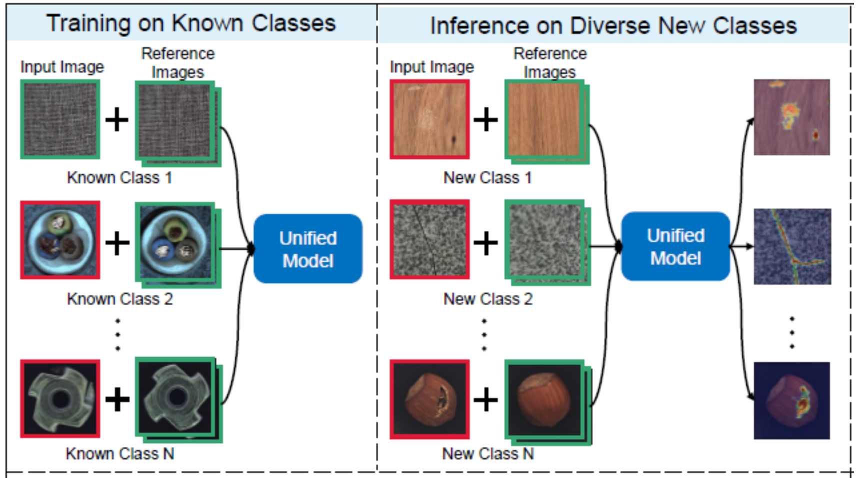
Xincheng Yao
Shanghai Jiao Tong University

Coauthors:
Zixin Chen, Chao Gao, Guangtao Zhai, Chongyang Zhang*

Preview

Class Generalizable Anomaly Detection

The objective is to train one **unified** AD model that can **generalize** to detect anomalies in diverse classes from **different domains** without **any retraining or fine-tuning** on the target data.

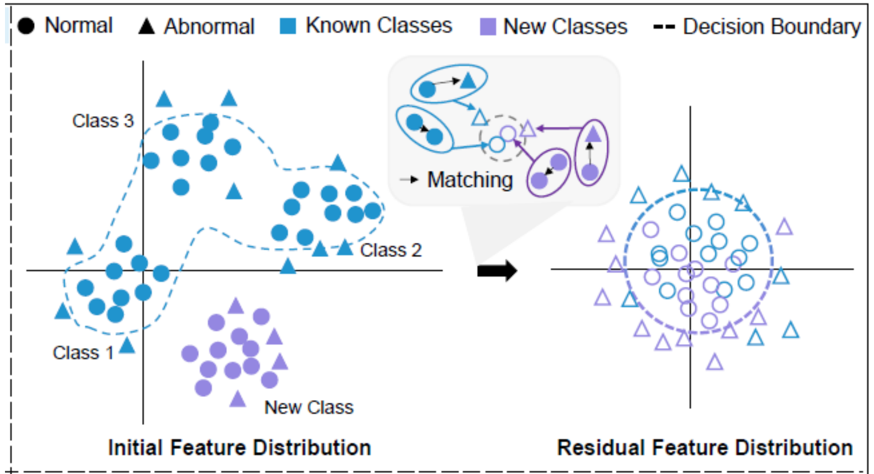


(a) Class-Generalizable Anomaly Detection

This AD task is challenging!

The main challenge is: *the normal patterns from different classes are significantly different.* This can lead to many misdetections of new classes.

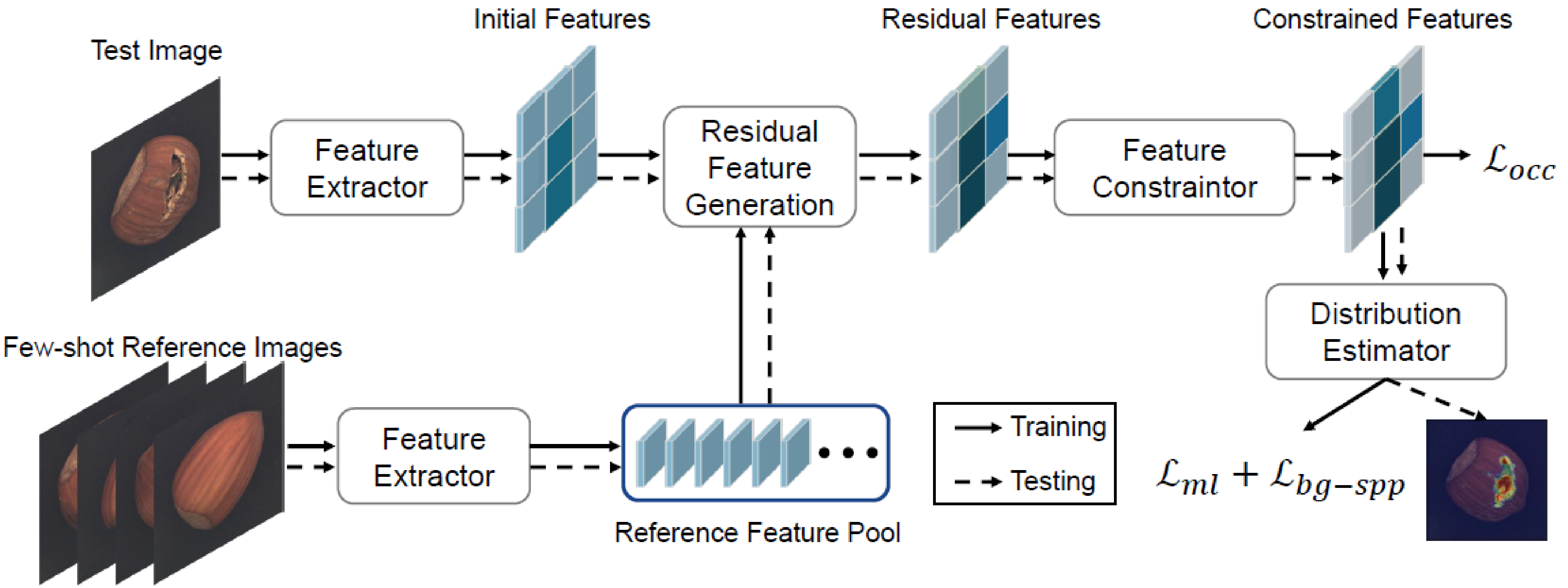
Our intuition: **residual features**!



(b) Residual Feature Learning

Preview

ResAD: A Simple But Effective Framework for Class Generalizable Anomaly Detection



Three parts: Residual Feature Generating, Feature Hypersphere Constraining, Feature Distribution Estimating.

Outline

NEURAL INFORMATION PROCESSING SYSTEMS

SHANGHAI JIAO TONG UNIVERSITY

1 Motivation

2 Our Approach: ResAD

3 Experiments

4 Ablations

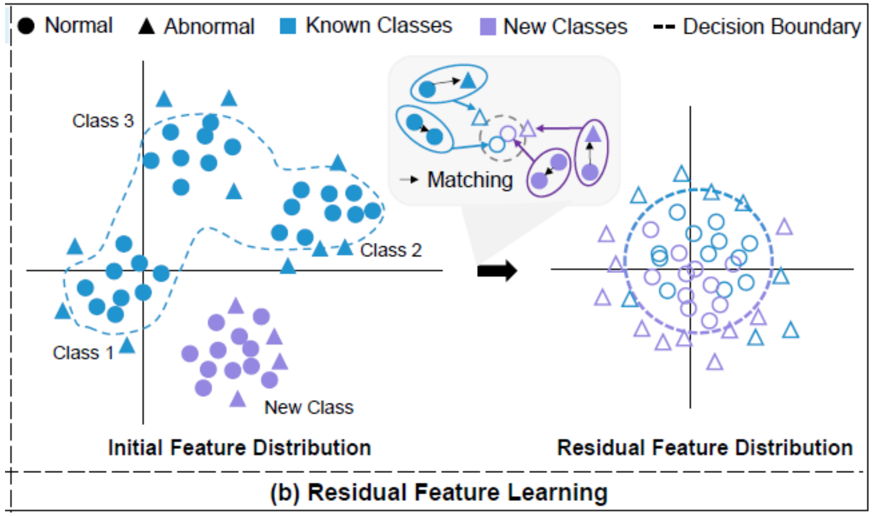
5 Conclusions

Motivation

Our core insight: Residual features are class-invariant representations!

Why previous AD models are not class generalizable?

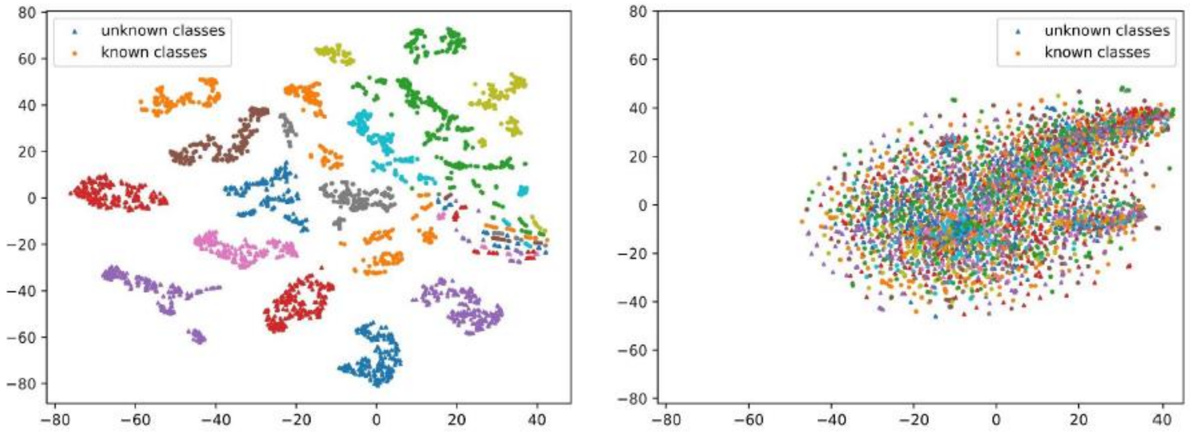
- The main challenge is: the normal patterns from different classes are significantly different (**class-variant representations**).
- Thus, normal patches from new classes may be mistaken as abnormal as they are quite different from the learned normal patterns.



(b) Residual Feature Learning

The characteristics of residual features:

- Residual features are **more class-invariant** compared to the significantly variant initial features.
- Residual features will be distributed in a **relatively fixed origin-centered region**, even in new classes.



(a) Initial Feature Distribution (b) Residual Feature Distribution

Motivation

NEURAL INFORMATION PROCESSING SYSTEMS

SHANGHAI JIAO TONG UNIVERSITY

Why can residual features be less sensitive to new classes compared to initial features?

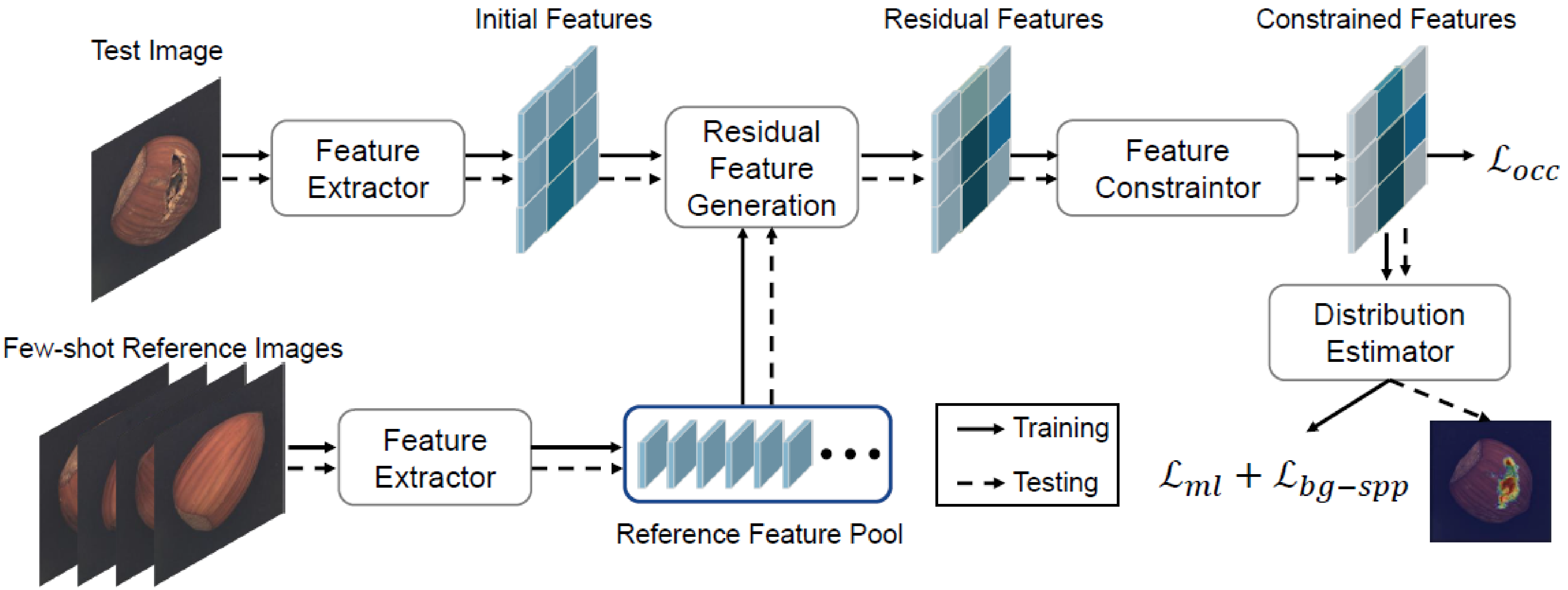
The definition of residual features:

Residual feature: $x^{l,r}_{h,w} = x^l_{h,w} - x^*_n$ Nearest Normal Reference Feature: $x^*_n = \mathrm{argmin}_{x \in \mathcal{P}_l} ||x - x^l_{h,w}||_2$

- Residual features are obtained by **matching** and then **subtracting**.
- From the principles of representation learning, we know that features of each class usually have some **class-related attributes** to the class for distinguishing from other classes.
- The ``class-related" means these attributes are **typical** to the class and **distinctive** from other classes.
- As class-related attributes can also exist in normal reference features, the matching process can be seen as **matching the most similar class-related attributes** to each input feature.
- By subtracting, the class-related components are very likely to be **mutually eliminated**.
- Thus, residual features will be distributed in an **origin-centered region**, even in new classes.

Our Approach: ResAD

- **ResAD, Model Overview:**



Three parts: Residual Feature Generating, Feature Hypersphere Constraining, Feature Distribution Estimating.

Our Approach: ResAD

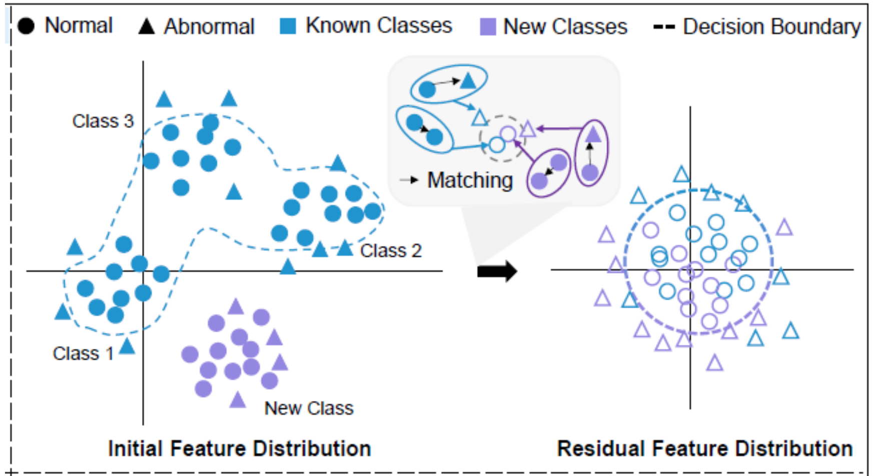
- **Residual Feature Generating**

- For each feature $x_{h,w}^l$, we will match it with the nearest normal reference feature from the corresponding reference feature pool, then convert to the residual feature by subtracting.

- **Reference Feature Pool**: The reference feature pools are utilized to store some normal features as reference for new classes.

- **Residual Feature:** We define the residual representation of $x_{h,w}^l$ to its closest normal reference feature as:

$$x_{h,w}^{l,r} = x_{h,w}^l - x_n^* \qquad x_n^* = \operatorname{argmin}_{x \in \mathcal{P}_l} ||x - x_{h,w}^l||_2$$



(b) Residual Feature Learning

Our Approach: ResAD

NEURAL INFORMATION PROCESSING SYSTEMS

SHANGHAI JIAO TONG UNIVERSITY

- **Feature Hypersphere Constraining**

- In order to **further reduce feature variations** and also **maintain the consistency in feature scales** among different classes, we propose a **Feature Constraintor** to constrain the initial normal residual features to a **spatial hypersphere**.
- **Abnormal Invariant OCC Loss**: we propose an abnormal invariant OCC loss to optimize our Feature Constraintor:

$$\mathcal{L}_{occ} = \frac{1}{L}\sum_{l=1}^{L}\left(\frac{1}{H_l W_l}\sum_{h=1}^{H_l}\sum_{w=1}^{W_l}(1-y_{h,w}^{l})||\sqrt{||x_{h,w}^{\prime,l,r}||_2+1}-1||_1 + y_{h,w}^{l}||x_{h,w}^{\prime,l,r}-x_{h,w}^{l,r}||_2\right)$$

- The loss can constrain the normal residual features to a hypersphere and keep abnormal residual features as **invariant** as possible.
- If we only constrain features to the hypersphere, the network may more easily overfit and simply map all features to the hypersphere.

Our Approach: ResAD

NEURAL INFORMATION PROCESSING SYSTEMS
SHANGHAI JIAO TONG UNIVERSITY

- **Feature Distribution Estimating**

- We employ the **normalizing flow** model as our **Feature Distribution Estimator** to estimate the residual feature distribution.
- The maximum likelihood loss function for learning normal residual feature distribution is as follows:

$$\mathcal{L}_{ml} = \frac{1}{L}\sum_{l=1}^{L}\left(\frac{1}{H_l W_l}\sum_{h=1}^{H_l}\sum_{w=1}^{W_l}\frac{C_l}{2}\log(2\pi) + \frac{1}{2}(z_{h,w}^l)^T z_{h,w}^l - \log|\det J_{h,w}^l|\right)$$

- In the class-generalizable AD task, it`s also valuable for us to effectively utilize abnormal samples that exist in known classes.
- Following BGAD[1], we employ the explicit boundary guided semi-push-pull loss to learn a more discriminative feature distribution estimator :

$$\mathcal{L}_{bg-spp} = \sum_{i=1}^{N_n}|\min(\log p_i - b_n, 0)| + \sum_{j=1}^{N_a}|\max(\log p_j - b_n + \tau, 0)|$$

[1] Xincheng Yao, Ruoqi Li, Jing Zhang, Jun Sun, and Chongyang Zhang. Explicit Boundary Guided Semi-Push-Pull Contrastive Learning for Supervised Anomaly Detection. CVPR, 2023.

Experiments

- **Datasets:**
  - Industrial AD datasets: MVTecAD, VisA, BTAD, MVTec3D.
  - Medical AD dataset: BraTS.
  - Video AD dataset: ShanghaiTech (we extract video frames as images for use).
- **Settings:**
  - We evaluate the **cross-dataset performance**.
  - We train on MVTecAD and test on other five datasets without any retraining.
  - For MVTecAD, we train AD models on VisA.
- **Metrics:**
  - Area under the curve of the receiver operating characteristic (AUROC), image-level and pixel-level.

Experiments

- **Comparison with Few-shot and CLIP-based AD methods:**

Setting	Datasets	Baselines RDAD CVPR2022	Baselines UniAD NeurIPS2022	Few-shot AD Methods (Non-CLIP-based) SPADE	PaDiM	PatchCore CVPR2022	RegAD ECCV2022	ResAD (Ours)	CLIP-based AD Methods WinCLIP CVPR2023	InCTRL CVPR2024	ResAD† (Ours)
	Industrial AD Datasets										
2-shot	MVTecAD	65.9/71.9	67.4/81.1	74.6/64.0	79.5/93.8	74.7/85.2	80.4/93.3	85.6/94.1	93.1/93.8	94.0/-	**94.4/95.6**
2-shot	VisA	56.4/79.9	52.1/81.8	71.7/65.4	68.7/91.5	65.0/80.4	70.6/93.3	79.9/**96.4**	81.9/94.9	**85.8/-**	84.5/95.1
2-shot	BTAD	82.7/87.3	67.1/85.6	80.7/65.4	88.9/95.2	80.9/83.1	87.2/93.9	**93.6/97.1**	85.5/95.8	92.3/-	91.1/96.4
2-shot	MVTec3D	58.7/90.4	51.7/89.4	62.5/78.6	59.6/94.3	58.8/83.4	59.5/96.4	64.5/95.4	74.1/96.8	68.9/-	**78.5/97.5**
2-shot	Average	65.9/82.4	59.6/84.5	72.4/68.4	74.2/93.7	69.8/83.0	74.4/94.2	80.9/95.8	83.7/95.3	85.3/-	**87.1/96.2**
	Medical AD Dataset										
2-shot	BraTS	49.8/66.7	59.5/88.5	58.0/92.8	49.4/90.2	58.2/93.5	54.6/81.4	65.7/91.2	55.9/91.5	**74.6/-**	67.9/**94.3**
	Video AD Dataset										
2-shot	ShanghaiTech	56.2/77.6	55.9/79.4	73.8/87.0	70.4/85.6	71.8/87.8	72.7/87.3	78.4/88.5	78.5/88.1	68.7/-	**82.4/91.9**
2-shot	**All Average**	61.6/79.0	58.9/84.3	70.2/75.6	69.4/91.8	68.2/85.6	70.8/90.9	78.0/93.8	78.2/93.5	80.8/-	**83.1/95.2**
	Industrial AD Datasets										
4-shot	MVTecAD	65.9/71.9	67.4/81.1	75.5/64.0	82.5/94.9	80.6/90.2	84.8/94.5	90.5/95.7	**94.6**/94.2	94.5/-	94.2/**96.9**
4-shot	VisA	56.4/79.9	52.1/81.8	75.0/65.4	75.3/93.3	71.7/87.1	78.0/93.5	86.2/97.4	84.1/95.2	87.7/-	**90.8/97.5**
4-shot	BTAD	82.7/87.3	67.1/85.6	81.7/65.5	89.9/95.8	84.0/89.4	90.8/94.9	**95.6/97.6**	87.2/95.8	91.7/-	91.5/96.8
4-shot	MVTec3D	58.7/90.4	51.7/89.4	62.3/78.6	62.8/94.5	61.5/87.1	62.3/96.7	70.9/97.3	76.0/97.0	69.1/-	**82.4/97.9**
4-shot	Average	65.9/82.4	59.6/84.5	73.6/68.4	77.6/94.6	74.5/88.5	79.0/94.9	85.8/97.0	85.5/95.6	85.8/-	**89.7/97.3**
	Medical AD Dataset										
4-shot	BraTS	49.8/66.7	59.5/88.5	66.3/94.8	60.6/94.5	71.2/95.9	60.0/87.3	74.7/94.0	67.3/93.2	76.9/-	**84.6/96.1**
	Video AD Dataset										
4-shot	ShanghaiTech	56.2/77.6	55.9/79.4	77.1/87.4	74.3/85.9	77.8/88.2	76.4/87.7	79.8/89.5	79.6/88.6	69.2/-	**84.3/92.6**
4-shot	**All Average**	61.6/79.0	58.9/84.3	73.0/76.0	74.2/93.2	74.5/89.7	75.4/92.4	83.0/95.3	81.5/94.0	81.5/-	**88.0/96.3**

- For new classes, the performance of conventional AD methods will drop dramatically (RDAD, UniAD).
- Our ResAD can achieve superior performance even without any re-modeling or fine-tuning.

Ablations

- **Ablation study results:**

(a) Framework ablation studies.

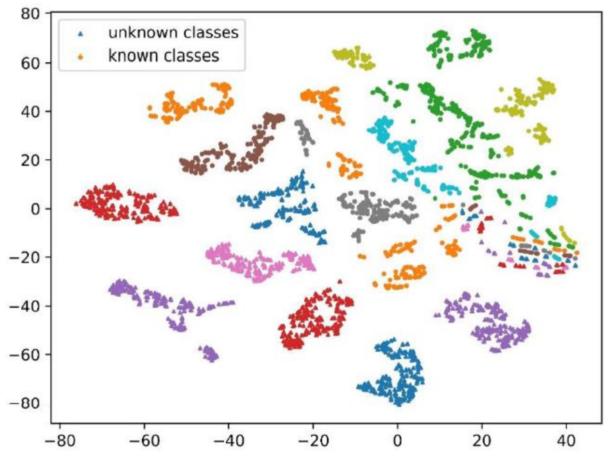
Model	I-AUROC	P-AUROC
Ours	90.5	95.7
w/o Residual Feature Learning	72.8	82.9
w/o Feature Constraintor	82.3	93.5
w/o Abnormal Invariant OCC Loss	84.9	93.9

Table 4: Anomaly detection and localization results when incorporating our method into UniAD. "RFL" represents residual feature learning.

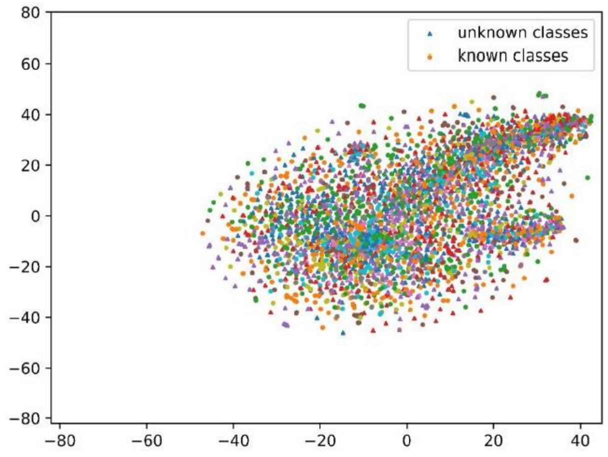
	MVTecAD	VisA	BTAD	MVTec3D
UniAD [47]	67.4/81.1	52.1/81.8	67.1/85.6	51.7/89.4
+ RFL (Ours)	93.0/94.9	72.7/86.1	87.3/94.0	76.7/96.9
Δ	+25.6/13.8	+20.4/3.3	+20.0/8.4	+25.0/7.0

- 1. Residual feature learning is of vital significance for class-generalizable anomaly detection.
- 2. Feature constraintor and abnormal invariant OCC loss are beneficial for achieving better cross-class performance.
- 3. Residual features can generalize to other AD models and significantly improve the models' class-generalizable capacity.

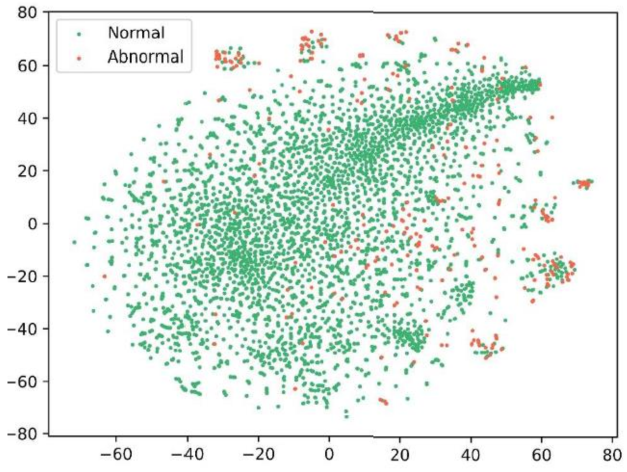
Visualization Results



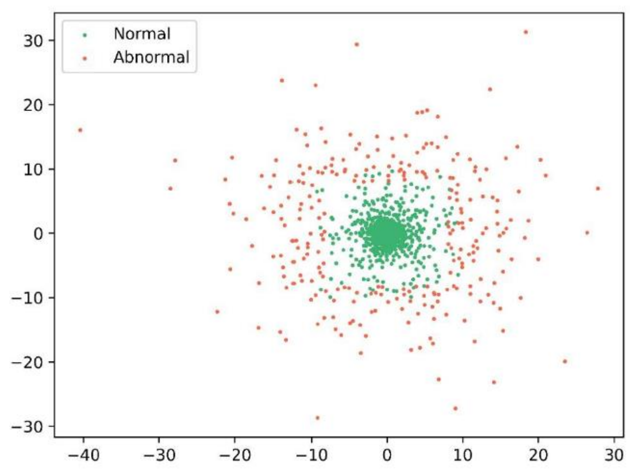
(a) Initial Feature Distribution



(b) Residual Feature Distribution



(c) Initial Residual Features



(d) Constrained Residual Features

Conclusions

NEURAL INFORMATION PROCESSING SYSTEMS

SHANGHAI JIAO TONG UNIVERSITY

Residual Features!

We conclude our finding for future research: **residual features are really effective for designing generalizable AD models**, and our feature constraining insight also has good reference values for future work.

SHANGHAI JIAO TONG UNIVERSITY

NEURAL INFORMATION PROCESSING SYSTEMS

Thanks!

Contact Us:
sunny_zhang@sjtu.edu.cn