Hierarchical Gaussian Mixture Normalizing Flow Modeling for Unified Anomaly Detection

ECCV 2024

Xincheng Yao
Shanghai Jiao Tong University

Oct 1 10:30-12:30

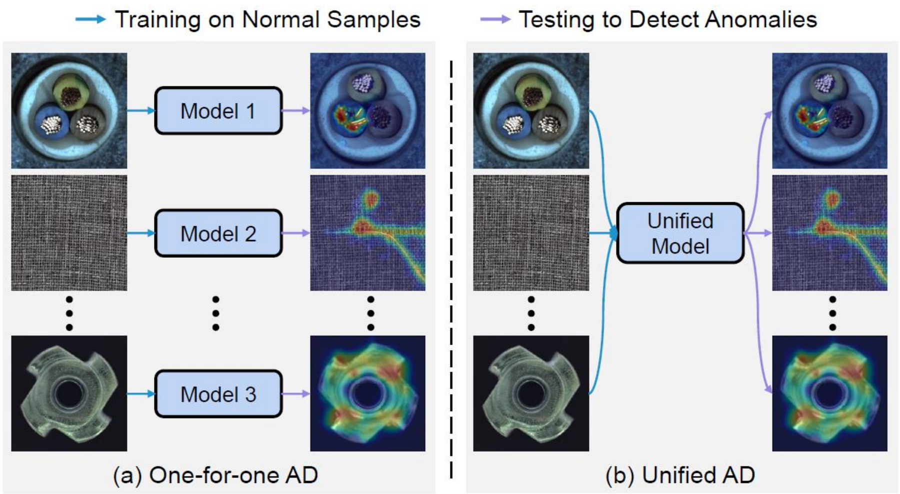
Coauthors:
Ruoqi Li, Zefeng Qian, Lu Wang, Chongyang Zhang*

Preview

Definition of Unified Anomaly Detection

One unified model is trained with normal samples from multiple classes.

The objective is to detect anomalies in multiple classes simultaneously.

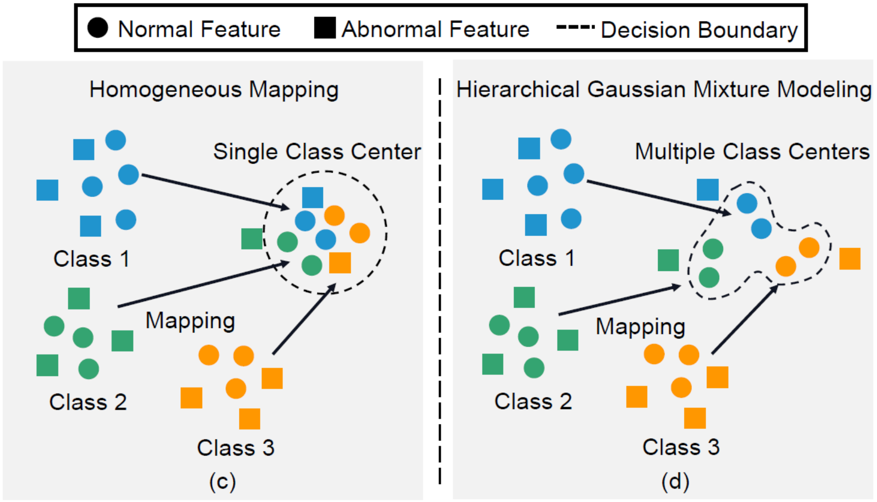


This AD task is challenging!

Previous research line: such as UniAD, addressing the "identical shortcut reconstruction" dilemma.

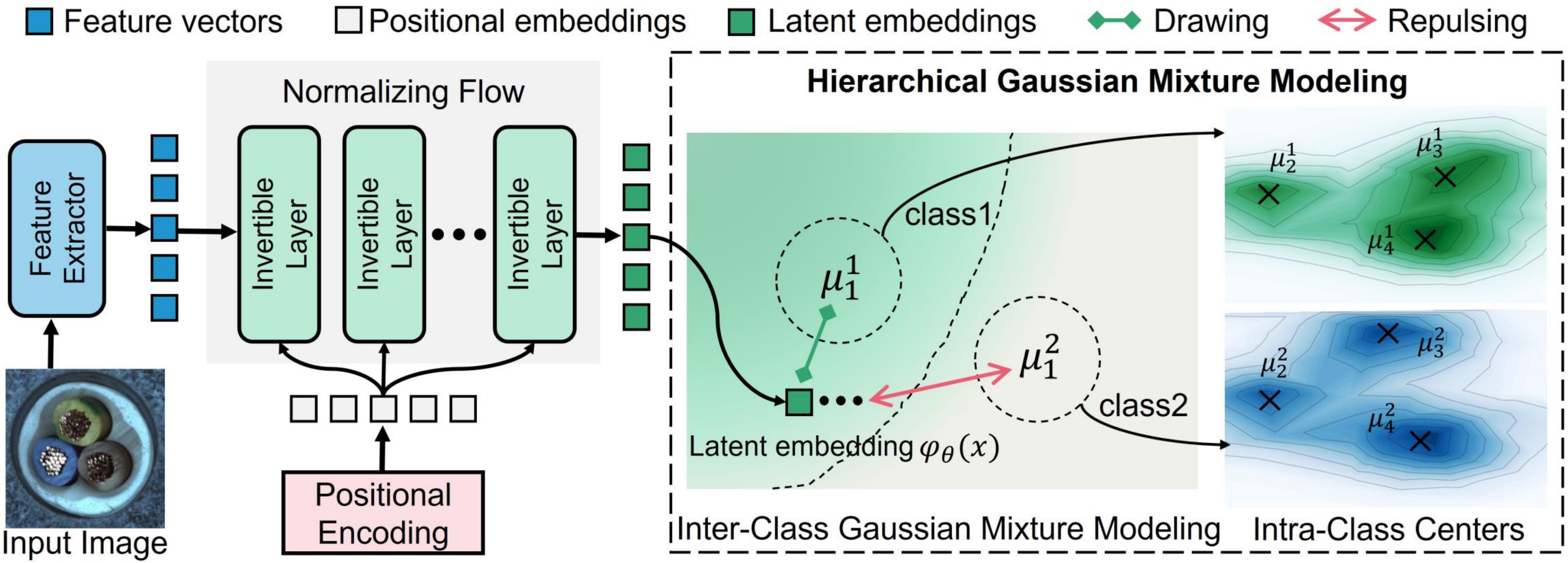
Our line: improving normalizing flow (NF) based AD methods to accomplish this task.

Popular NF-based AD methods may fall into a "**homogeneous mapping**" issue.



Preview

HGAD: Hierarchical Gaussian Mixture Modeling for Unified Anomaly Detection



Three parts: a feature extractor, a normalizing flow model, and the hierarchical Gaussian mixture modeling.

Outline

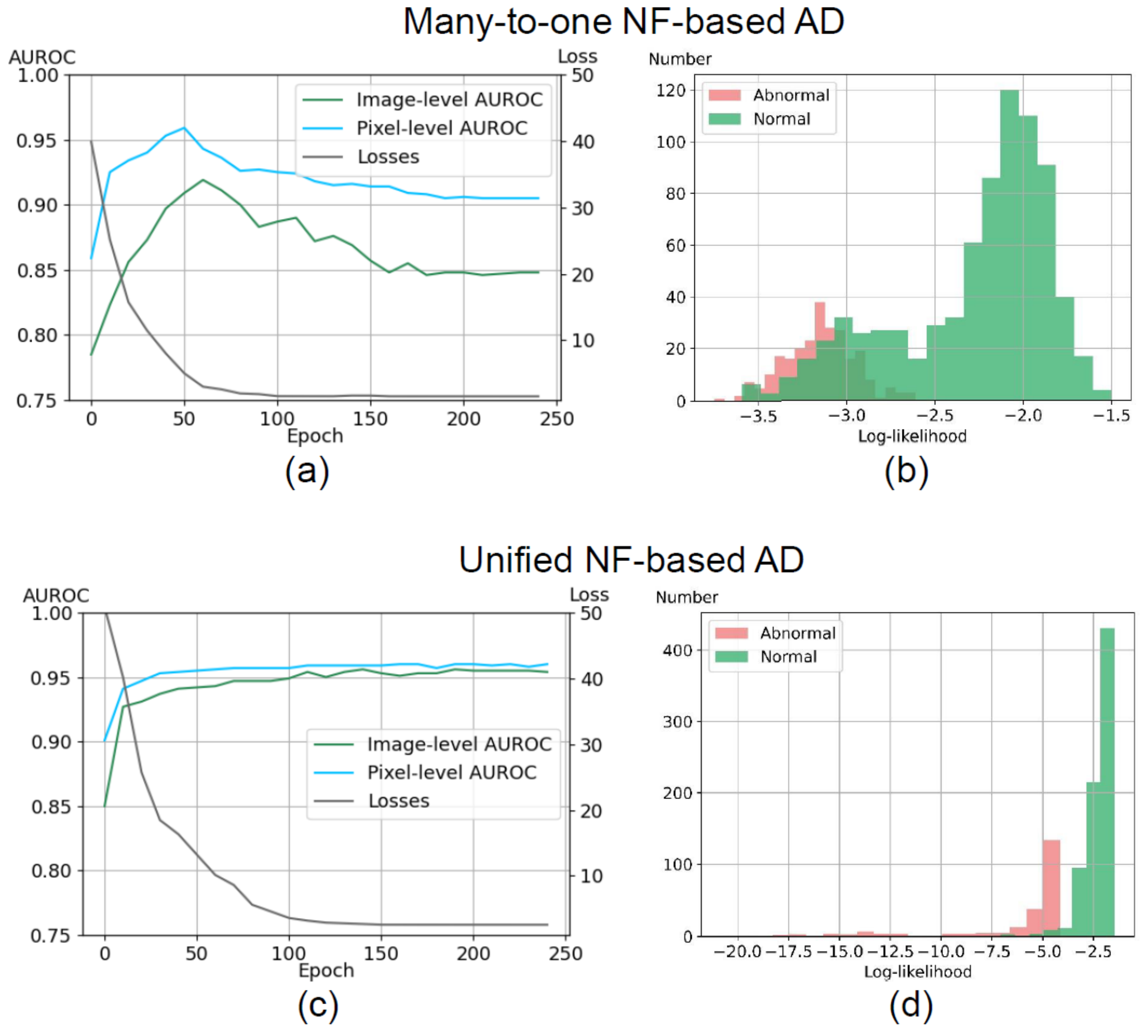
SHANGHAI JIAO TONG UNIVERSITY

Motivation

Revisiting Normalizing Flow Based Anomaly Detection Methods

"Homogeneous mapping" issue:

- When applying previous NF-based AD model (e.g., FastFlow) to the unified AD task, the performance drops severely while the losses continue going extremely small.
- **Homogeneous mapping**: the NF model may map all inputs to much close latent variables, both normal and abnormal features have large log-likelihoods.



Motivation

An explanation from the NF model structure perspective

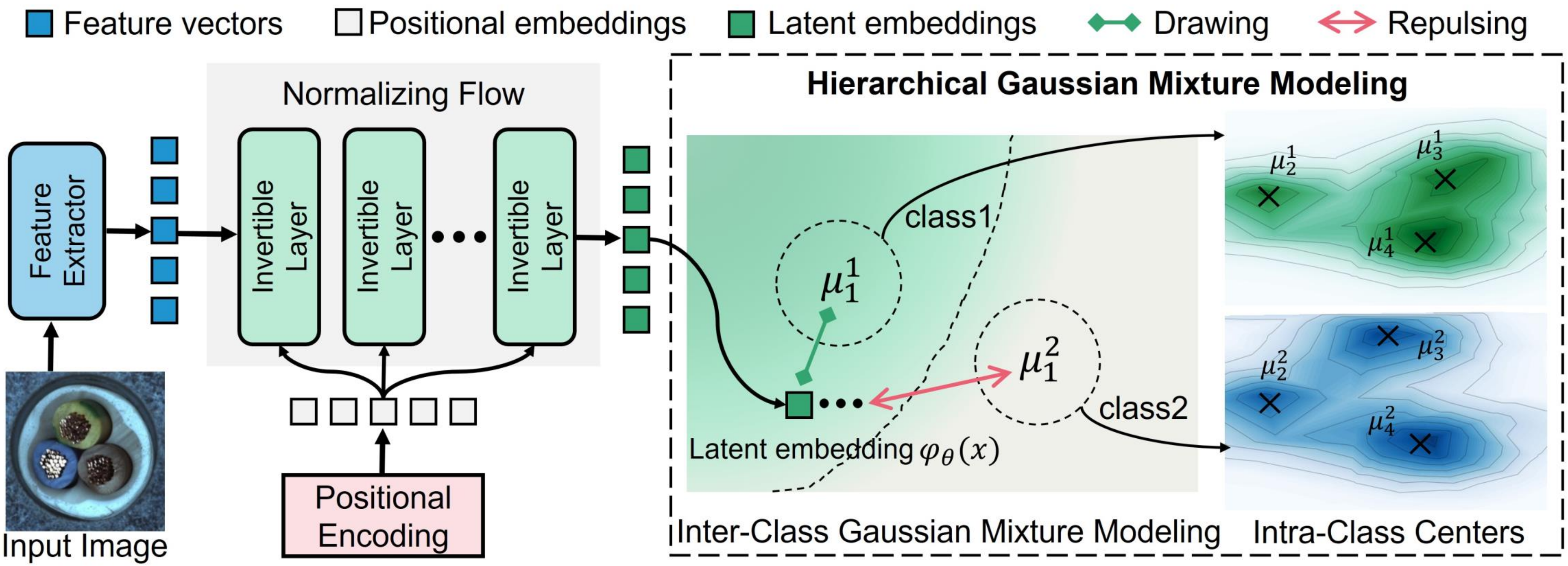
The formula of the forward affine coupling in NF model is：

$$x_1, x_2 = \mathrm{split}(x_n)\ ;\ z_1 = x_1, z_2 = x_2 \odot \exp(s(x_1)) + t(x_1)\ ;\ z = \mathrm{cat}(z_1, z_2)$$

- $s(x_1)$ and t(x_1) are transformation coefficients predicted by a learnable neural network.
- With the maximum likelihood loss pushing all z to fit $\mathcal{N}(0, \mathbb{I})$ (in previous NF-based AD methods), the model has no need to distinguish different class features.
- It is more likely to take a bias to predict all $s(\cdot)$ **to be very small negative numbers (**$\rightarrow -\infty$**)** and **t**$(\cdot)$ **close to zero**.
- The impact is that the model could also **fit x_a to $\mathcal{N}(0, \mathbb{I})$ well with the bias**.
- **If we map different class features to different class centers, the model is harder to simply take a bias solution**.

Our Approach: HGAD

- **HGAD, Model Overview:**



Three parts: a feature extractor, a normalizing flow model, and the hierarchical Gaussian mixture modeling.

Our Approach: HGAD

- **Inter-Class Gaussian Mixture Modeling**

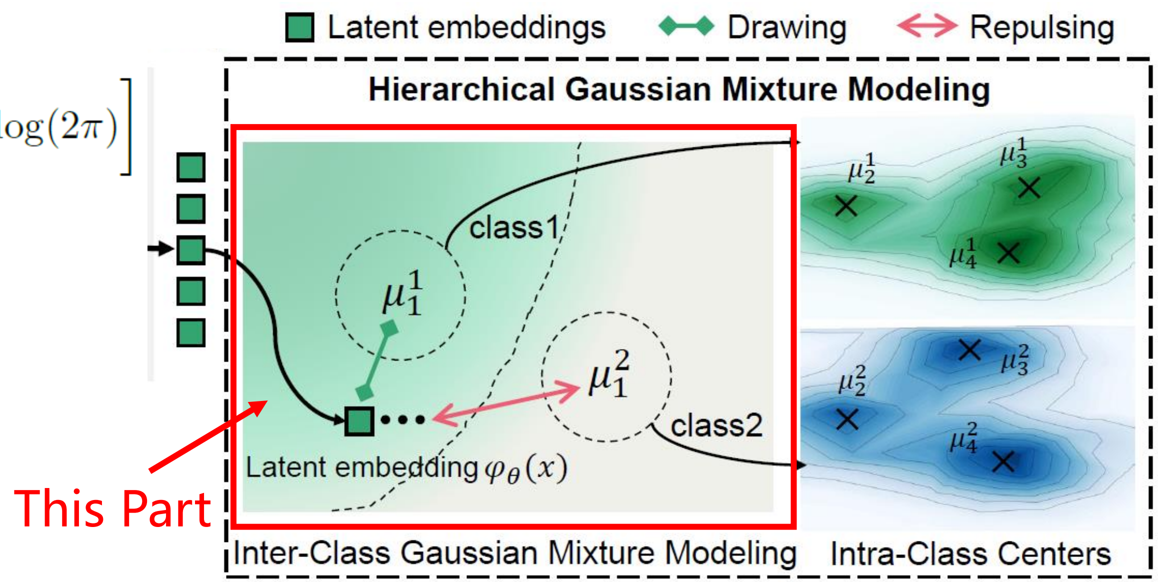
- To better fit the complex multi-class distribution, we propose the inter-class Gaussian mixture modeling approach.
- A Gaussian mixture model with class-dependent means μ_y and covariances Σ_y is used as the prior distribution in the latent space:

$$p_{Z|Y}(z|y) = \mathcal{N}(z; \mu_y, \Sigma_y) \quad \text{and} \quad p_Z(z) = \sum_y p(y)\mathcal{N}(z; \mu_y, \Sigma_y)$$

- We parameterize the class weights $p(Y)$ through a learnable vector ψ, with $p(y) = \text{softmax}_y(\psi)$.
- **The optimization loss function is**:

$$\mathcal{L}_g = \mathbb{E}_{x\sim p(X)}\left[-\underset{y}{\text{logsumexp}}\left(-\frac{||\varphi_\theta(x)-\mu_y||_2^2}{2}+c_y\right)-\log|\det J|+\frac{d}{2}\log(2\pi)\right]$$

- c_y denotes logarithmic class weights:
- $c_y := \log p(y) = \text{logsoftmax}_y(\psi)$.



Our Approach: HGAD

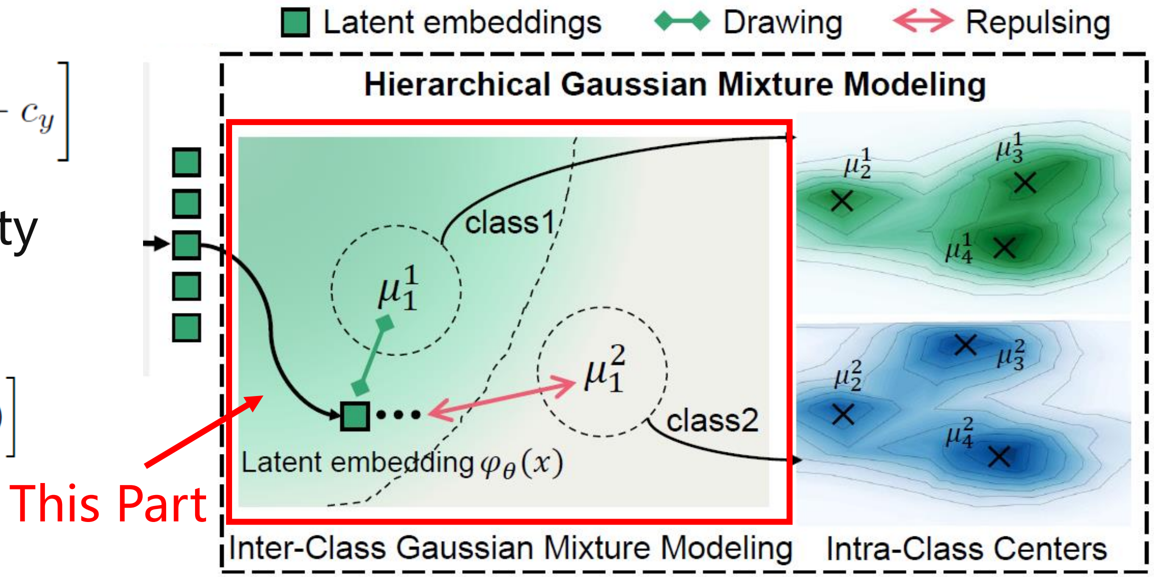
- **Mutual Information Maximization**

- The inter-class Gaussian modeling can only ensure the latent features are drawn together to the whole distribution.
- The last loss function only has the drawing characteristic, but without the repulsion property for separating different classes.
- **This may cause different class centers to collapse into the same center.**
- We propose a **mutual information maximization loss** to introduce the class repulsion property:

$$\mathcal{L}_{mi} = -\mathbb{E}_{(x,y)\sim p(X,Y)}\left[\underset{y}{\mathrm{logsoftmax}}\left(-\frac{||\varphi_\theta(x)-\mu_{y'}||_2^2}{2}+c_{y'}\right)-c_y\right]$$

- We can also introduce the class repulsion property by minimizing the inter-class entropy:

$$\mathcal{L}_e = \mathbb{E}_{x\sim p(X)}\left[\sum_y -\underset{y}{\mathrm{softmax}}(-||\varphi_\theta(x)-\mu_{y'}||_2^2/2)\cdot\underset{y}{\mathrm{logsoftmax}}(-||\varphi_\theta(x)-\mu_{y'}||_2^2/2)\right]$$



Our Approach: HGAD

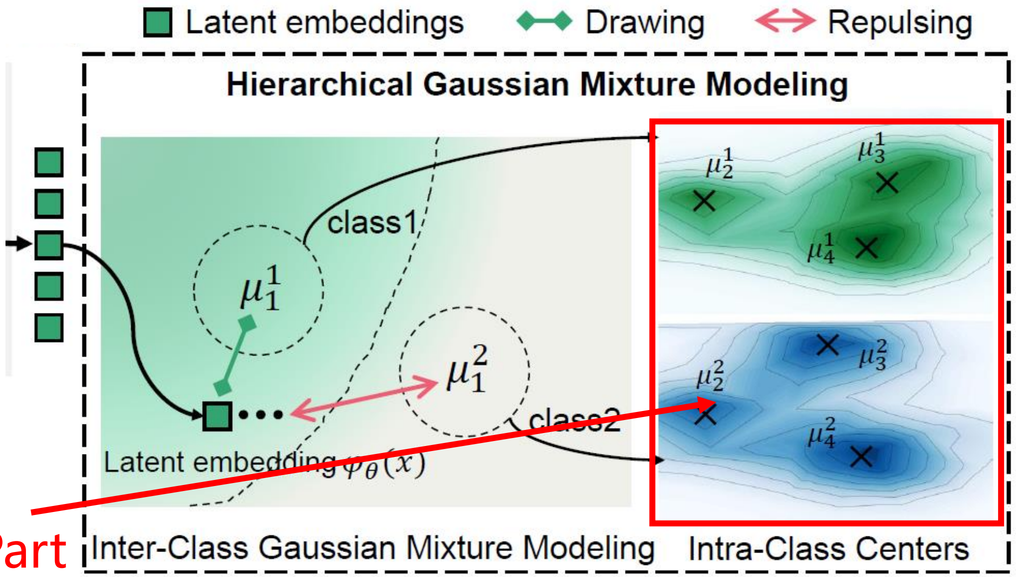
- **Learning Intra-Class Mixed Class Centers**

- In real-world scenarios, even one object class may contain diverse normal patterns.
- To better model intra-class distribution, we further extend the Gaussian prior $p(Z|y) = \mathcal{N}(\mu_y, \Sigma_y)$ to mixture Gaussian prior $p(Z|y) = \sum_{i=1}^{M} p_i(y)\mathcal{N}(\mu_i^y, \Sigma_i^y)$.
- To avoid numerically ill-defined loss, we decouple the inter-class Gaussian mixture modeling and the intra-class latent centers learning.
- For each class y, we learn a main class center μ_1^y and the delta vectors $\{\Delta\mu_i^y\}_{i=1}^{M}$($\Delta\mu_1^y$ is fixed to 0).
- **The loss function for intra-class centers learning is:**

$$\mathcal{L}_{in} = \mathbb{E}_{(x,y)\sim p(X,Y)}\left[-\underset{i}{\text{logsumexp}}\left(-\frac{\|\varphi_\theta(x) - (SG[\mu_1^y] + \Delta\mu_i^y)\|_2^2}{2} + c_i^y\right) - \log|\det J|\right]$$

- $SG[\cdot]$ means to stop gradient backpropagation.
- Combining the three parts, we form a hierarchical Gaussian mixture normalizing flow modeling method

This Part



Experiments

- **Datasets:**
  - MVTecAD: 5534 high-resolution images, 15 product categories, 73 anomaly types, and 1900 abnormal regions.
  - VisA: This dataset contains 10821 images with 9621 normal and 1200 anomalous samples, total 12 product categories.
  - BTAD: This dataset contains 2830 real-world images of 3 industrial products.
  - MVTec3D-RGB: This dataset contains 4147 RGB images from 10 real-world categories.
- **Metrics:**
  - Area under the curve of the receiver operating characteristic (AUROC), image-level and pixel-level.

Experiments

- **Detailed Results on MVTecAD:**

Category	Baseline Methods			Unified Methods			Normalizing Flow Based Methods		
	PaDiM	MKD	DRAEM	PMAD	UniAD	OmniAL	FastFlow	CFLOW	HGAD (Ours)
Carpet	93.8/97.6	69.8/95.5	98.0/98.6	99.0/97.9	99.8/98.5	98.7/**99.4**	91.6/96.7	98.8/97.5	**100**±0.00/**99.4**±0.05
Grid	73.9/71.0	83.8/82.3	99.3/98.7	96.2/95.6	98.2/96.5	**99.9**/**99.4**	85.7/96.8	95.9/94.1	99.6±0.09/99.1±0.08
Leather	99.9/84.8	93.6/96.7	98.7/97.3	**100**/99.2	**100**/98.8	99.0/99.3	93.7/98.2	**100**/98.1	**100**±0.00/**99.6**±0.00
Tile	93.3/80.5	89.5/85.3	99.8/98.0	99.8/94.5	99.3/91.8	99.6/**99.0**	99.2/95.8	97.9/92.2	**100**±0.00/96.1±0.09
Wood	98.4/89.1	93.4/80.5	**99.8**/96.0	99.6/89.0	98.6/93.2	93.2/**97.4**	98.0/92.0	99.0/92.7	99.5±0.08/95.9±0.09
Bottle	97.9/96.1	98.7/91.8	97.5/87.6	99.8/98.4	99.7/98.1	**100**/**99.2**	**100**/94.0	98.7/96.4	**100**±0.00/98.6±0.08
Cable	70.9/81.0	78.2/89.3	57.8/71.3	93.5/95.4	95.2/**97.3**	**98.2**/**97.3**	90.9/95.2	80.4/92.9	97.3±0.26/95.2±0.49
Capsule	73.4/96.9	68.3/88.3	65.3/50.5	80.5/97.0	86.9/98.5	95.2/96.9	90.5/98.6	75.5/97.7	**99.0**±0.40/**99.2**±0.05
Hazelnut	85.5/96.3	97.1/91.2	93.7/96.9	99.6/97.4	99.8/98.1	95.6/98.4	98.9/96.6	97.1/95.7	**99.9**±0.08/**98.8**±0.05
Metal nut	88.0/84.8	64.9/64.2	72.8/62.2	98.0/91.7	99.2/94.8	99.2/**99.1**	96.5/97.2	87.8/84.4	**100**±0.00/97.8±0.29
Pill	68.8/87.7	79.7/69.7	82.2/94.4	89.4/93.4	93.7/95.0	**97.2**/**98.9**	90.4/96.1	88.0/90.7	96.3±0.73/98.8±0.05
Screw	56.9/94.1	75.6/92.1	92.0/95.5	73.3/96.6	87.5/98.3	88.0/98.0	76.8/95.9	59.5/93.9	**95.5**±0.16/**99.3**±0.12
Toothbrush	95.3/95.6	75.3/88.9	90.6/97.7	95.8/98.2	94.2/98.4	**100**/**99.4**	86.1/97.1	78.0/95.7	91.2±0.37/99.1±0.05
Transistor	86.6/92.3	73.4/71.7	74.8/64.5	97.2/93.3	**99.8**/**97.9**	93.8/93.3	85.7/93.8	86.7/92.3	97.7±0.21/91.9±0.26
Zipper	79.7/94.8	87.4/86.1	98.8/98.3	96.0/96.1	95.8/96.8	**100**/**99.5**	93.8/95.7	92.2/95.7	**100**±0.04/99.0±0.09
Mean	84.2/89.5	81.9/84.9	88.1/87.2	94.5/95.6	96.5/96.8	97.2/**98.3**	91.8/96.0	89.0/94.0	**98.4**±0.08/97.9±0.05

- Compared with the one-for-one NF-based AD counterparts, our method can improve the unified AD performance significantly.
- Our HGAD also surpasses the SOTA unified AD methods, UniAD and OmniAL, demonstrating our superiority.

Experiments

- **More Results on Other Datasets:**

Dataset	PaDiM	MKD	DRAEM	PMAD	UniAD	OmniAL	FastFlow	CFLOW	HGAD (Ours)
BTAD	93.8/96.6	89.7/96.2	91.2/91.9	93.8/97.3	94.0/97.2	-/-	92.9/95.3	93.0/96.6	**94.9**±0.08/**98.0**±0.10
MVTec3D-RGB	77.4/96.3	73.5/95.9	73.9/95.5	75.4/95.3	77.5/96.6	-/-	67.9/90.2	71.6/95.7	**87.1**±0.16/**97.7**±0.05
VisA	86.8/97.0	74.2/93.9	85.5/90.5	-/-	92.8/98.1	87.8/96.6	77.2/95.1	88.0/95.9	**97.1**±0.09/**98.9**±0.05
Union	79.0/91.4	72.1/88.9	66.4/82.7	-/-	86.9/95.5	-/-	57.2/78.8	55.7/82.9	**93.5**±0.23/**97.5**±0.14

- We can outperform the best competitors also on BTAD, MVTec3D-RGB, and VisA datasets, especially on MVTec3D-RGB and VisA .
- To more sufficiently evaluate the unified AD performance, we combine all four datasets to form a 40-class dataset, Union.

Ablations

- **Ablation study results:**

(a) Hierarchical Gaussian mixture prior.

SGC	FMC	**ICG**	**MIM**	**Intra**	Det.	Loc.
✓	-	-	-	-	89.0	94.0
-	✓	-	-	-	93.1	95.8
-	-	✓	-	-	94.5	96.7
-	-	✓	✓	-	96.3	96.6
-	✓	-	-	✓	96.3	97.1
-	-	✓	✓	✓	**97.7**	**97.6**

(b) Number of intra-class centers.

# Centers	Det.	Loc.
3	97.5	97.3
5	97.6	97.1
10	**97.7**	**97.6**
15	97.6	97.3
20	97.4	97.2

(c) Anomaly criterion.

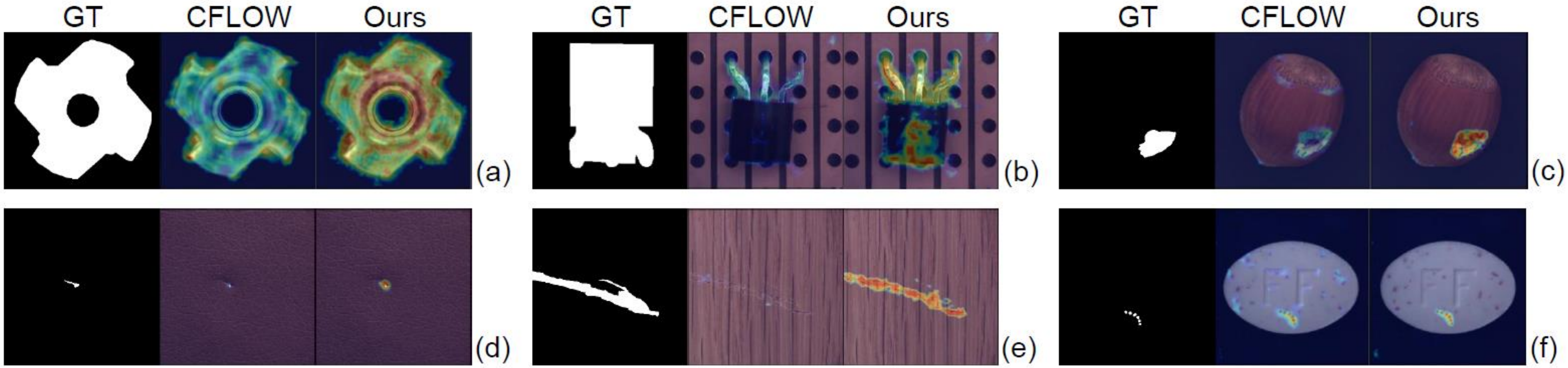
Logps	**Entropy**	Det.	Loc.
✓	-	94.4	97.1
-	✓	96.6	97.0
✓	✓	**97.7**	**97.6**

(d) Hyperparameters.

λ_1	λ_2	Det.	Loc.	λ_1	λ_2	Det.	Loc.
1	1	96.5	96.7	0.5	100	98.4	97.8
1	5	97.6	97.3	**1**	**100**	**98.4**	**97.9**
1	10	97.7	97.6	5	100	98.3	97.9
1	50	98.2	97.8	10	100	98.3	97.8
1	**100**	**98.4**	**97.9**	20	100	97.8	97.6

- 1、Mapping to multiple class centers is effective, and our inter-class Gaussian mixture modeling is can further improve the results compared to plain fixed multiple centers.
- 2、Mutual information maximization can bring promotion by 1.8% for image-level detection.
- 3、Learning intra-class mixed class centers could bring an increase of 1.6% for detection and 1.0% for localization, respectively.
- 4、Combining all these, the hierarchical Gaussian mixture modeling method can achieve the best results.

Qualitative Results



(a) and (b) both represent global anomalies,
(c) contains large cracks, (d) shows small dints,
(e) contains texture scratches, and (f) shows color anomalies.

Conclusions

Hierarchical Gaussian Mixture Normalizing Flow Modeling for Unified Anomaly Detection

For challenging **unified anomaly detection,** NF-based AD methods may fall into a "**homogeneous mapping**" issue

To address this, we propose a novel HGAD against learning the bias with three key improvements: **inter-class Gaussian mixture modeling, mutual information maximization, and intra-class mixed class centers learning strategy**

Thanks!

Contact Us:
sunny_zhang@sjtu.edu.cn