Deep Learning for Medical Image Analysis

COMP5423

Hao CHEN

Dept. of CSE,CBE&LIFS, HKUST

jhc@cse.ust.hk

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

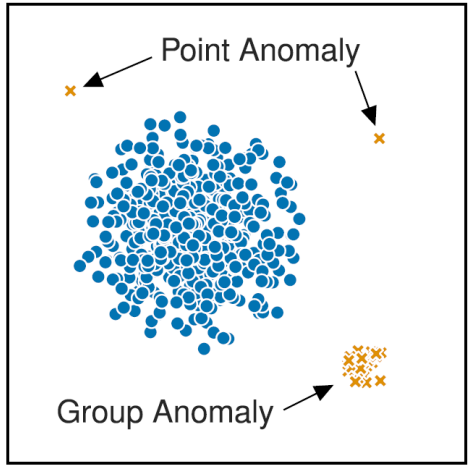
Anomaly Detection in MIA

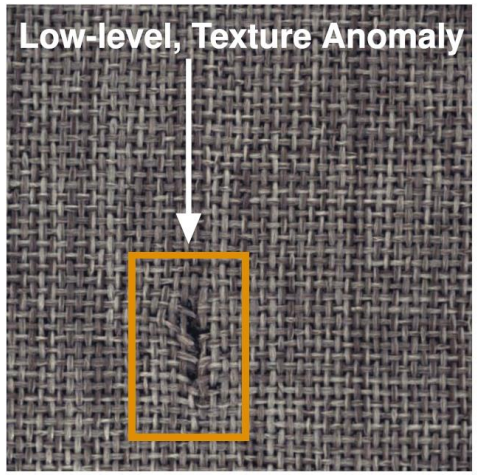
- Introduction
- Reconstruction-based methods
- Ensemble-based methods
- Self-supervised methods
- Non-OCC settings
- Challenge and future direction

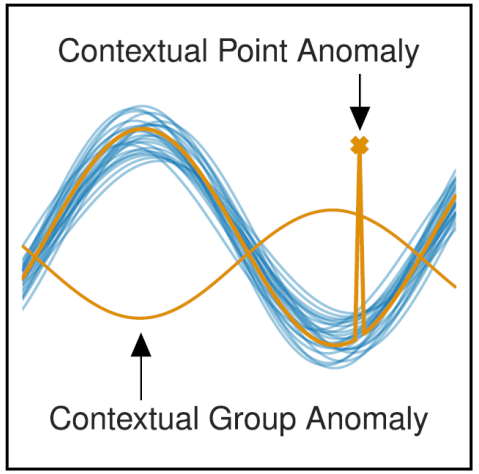
Introduction

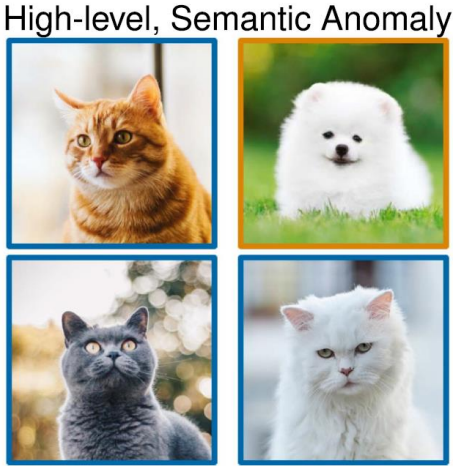
- Definition

An *anomaly* is an observation that deviates considerably from normality. Also known as outlier/novelty.





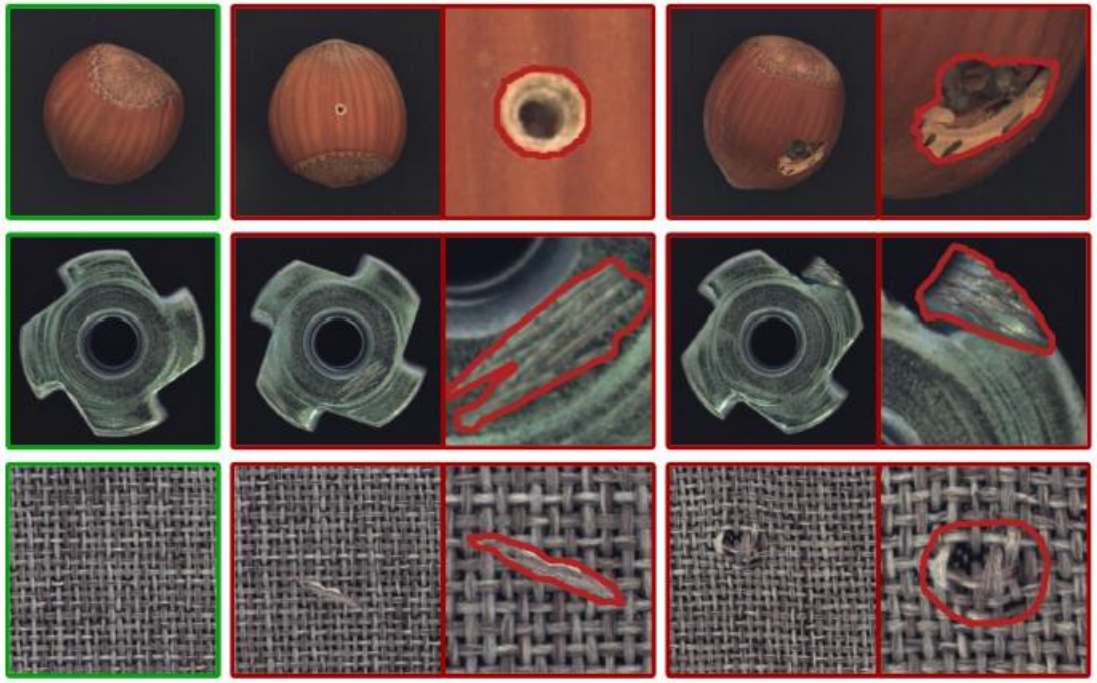




Ruff, Lukas, et al. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE (2021).

Introduction

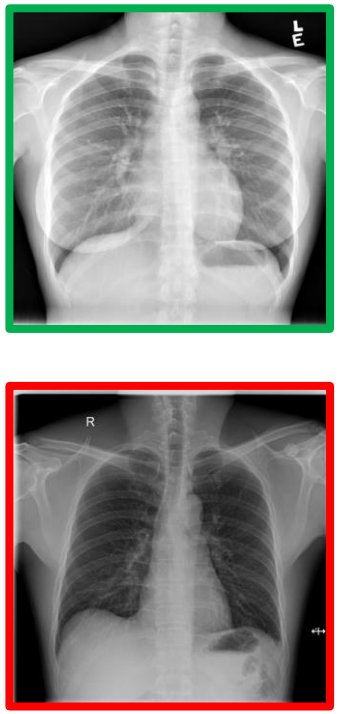
- Examples in industrial scenarios



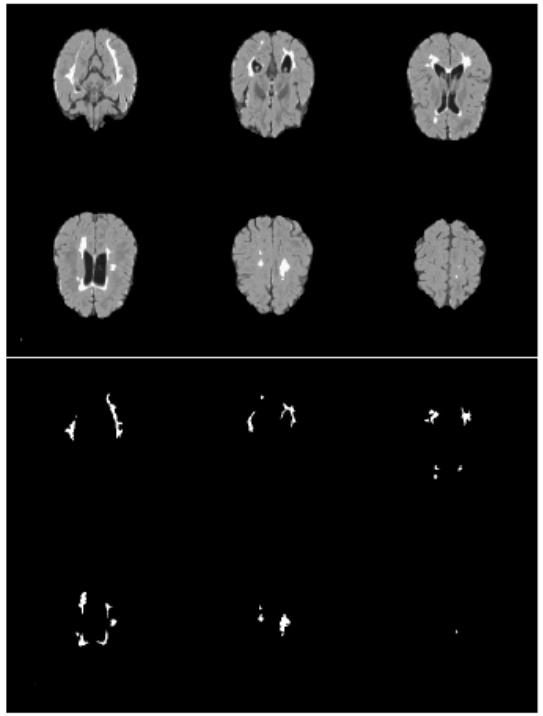
Bergmann, Paul, et al. MVTec AD--A comprehensive real-world dataset for unsupervised anomaly detection. CVPR 2019.

Introduction

- Examples in medical scenarios



Chest X-ray



Brain MRI

Introduction

- Anomaly Detection

Supervised learning of every possible pathology is unrealistic for many primary care applications like health screening.

Due to the difficulty of obtaining labeled anomalous data, most of Anomaly Detection methods are “unsupervised”:

Learning a model of normality from *only normal data* so that anomalies become detectable through deviations from the model.

Also known as the *One-class Classification (OCC).*

Ruff, Lukas, et al. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE (2021).

Introduction

Evaluation Metrics - Image-level anomaly detection

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = TPR = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

$$Accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

Threshold-independent metrics AUROC and AUPRC are also widely used.

Introduction

Evaluation Metrics - Pixel-level anomaly segmentation

$$Dice = \frac{2|\widehat{M} \cap M|}{|\widehat{M}| + |M|}$$

where $\widehat{M}$ denotes the predicted anomaly mask and M denotes the ground truth.

Anomaly Detection in MIA

- Introduction
- Reconstruction-based methods
- Ensemble-based methods
- Self-supervised methods
- Non-OCC settings
- Challenge and future direction

Reconstruction-based methods

- Reconstruction-based methods learn a model that is optimized to well-reconstruct normal data instances, thereby aiming to detect anomalies by failing to accurately reconstruct them under the learned model.

Reconstruction-based methods

- Formulation – Training

$\{x_i \in \mathcal{X}\ (i = 1, \ldots, n)\}$ are normal images.

The training objective is

$$\min_{\theta} \frac{1}{n} \sum_{i=1}^{n} \|\boldsymbol{x}_i - (\phi_d \circ \phi_e)_\theta(\boldsymbol{x}_i)\|^2$$

where ϕ_e is the encoder that maps the image to latent vector z_i

and ϕ_d is the decoder that maps the latent vector to reconstruction $\hat{x}_i$

Ruff, Lukas, et al. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE (2021).

Reconstruction-based methods

- Formulation – Testing

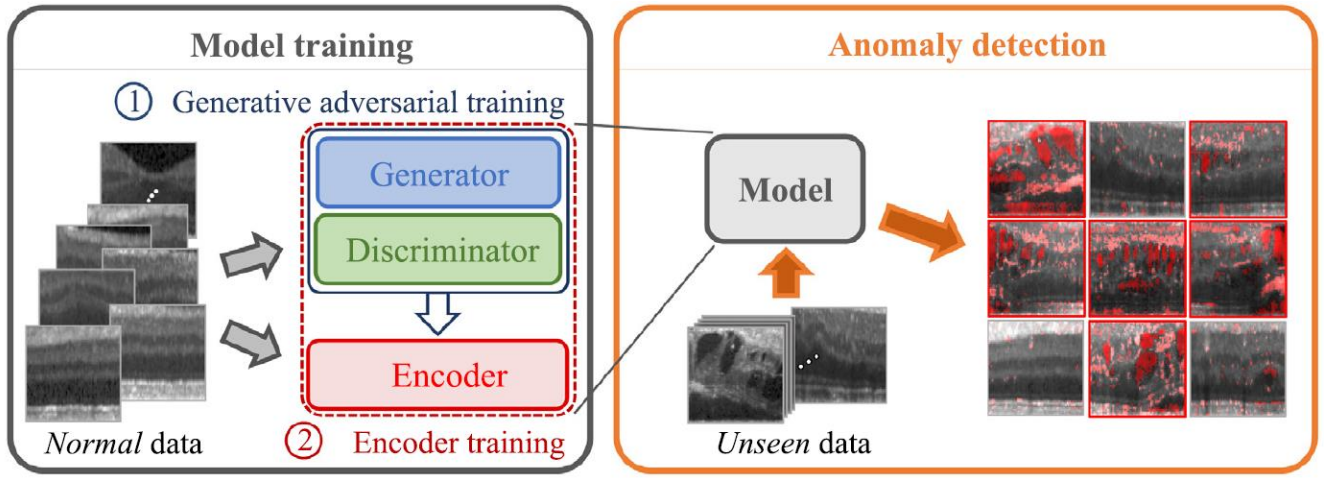
The anomaly score is usually defined by the reconstruction error:

$$s(\boldsymbol{x}) = \left\| \boldsymbol{x} - (\phi_d \circ \phi_e)_\theta(\boldsymbol{x}) \right\|^2$$

Ruff, Lukas, et al. A unifying review of deep and shallow anomaly detection. Proceedings of the IEEE (2021).

Reconstruction-based methods

- f-AnoGAN



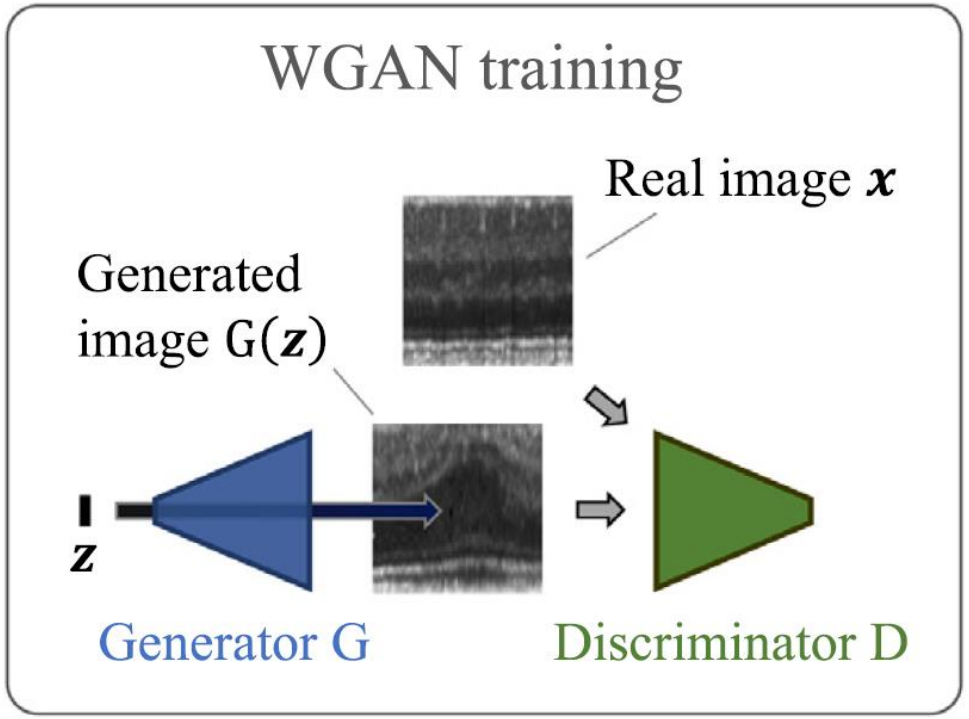
Step1: Build a generative model of normal data.

Step2: Train an encoder to map the image to GAN's latent space.

Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.

Reconstruction-based methods

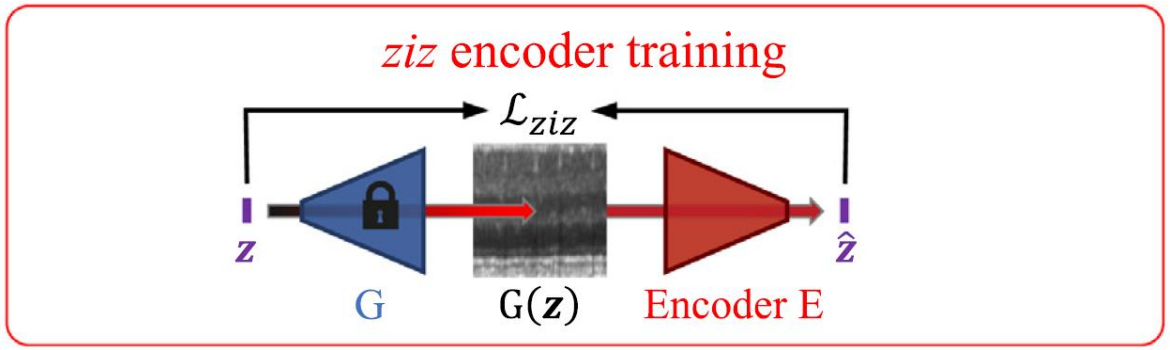
- Wasserstein GAN[2] (WGAN) training



[1] Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.
[2] Arjovsky, et al. Wasserstein generative adversarial networks. ICML 2017.

Reconstruction-based methods

- Three strategies for encoder training - ***ziz***

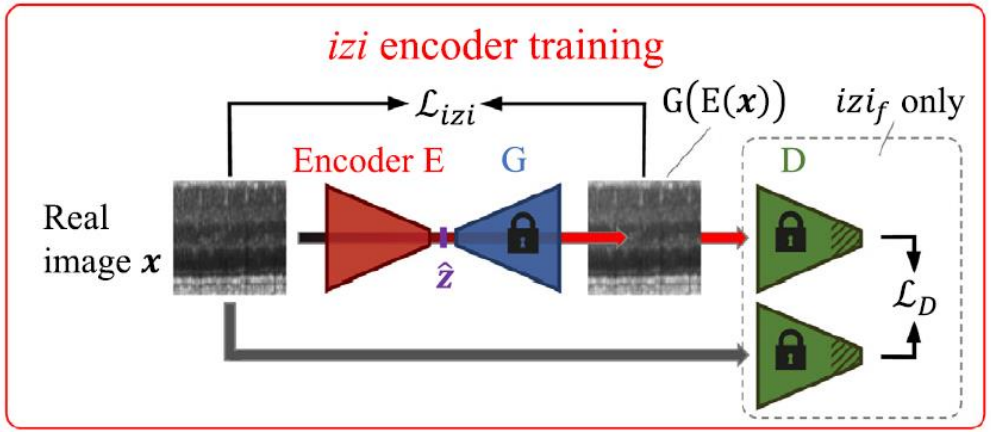


$$\mathcal{L}_{ziz}(\mathbf{z}) = \frac{1}{d}\|\mathbf{z} - E\big(G(\mathbf{z})\big)\|^2$$

Drawback: the encoder only “sees” generated images but never receives real input images.

Reconstruction-based methods

- Three strategies for encoder training - ***izi***



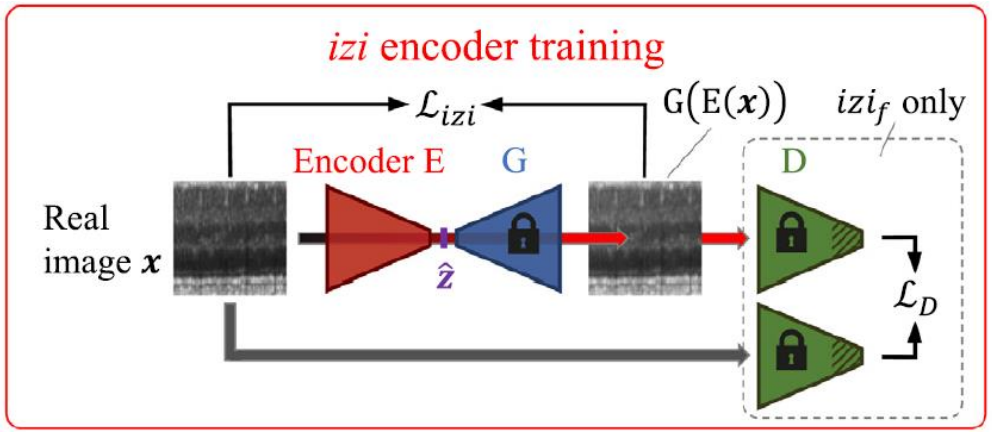
$$\mathcal{L}_{izi}(\mathbf{x}) = \frac{1}{n}\|\mathbf{x} - G\big(E(\mathbf{x})\big)\|^2$$

Drawback: Since the true target location in the z-space of a given query image is unknown, we can only indirectly measure the accuracy of the image to z mapping through mapping back to the image space and computing the image-to-image residual.

Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.

Reconstruction-based methods

- Three strategies for encoder training - $\boldsymbol{izi_f}$



Improvement: Discriminator guided izi encoder training

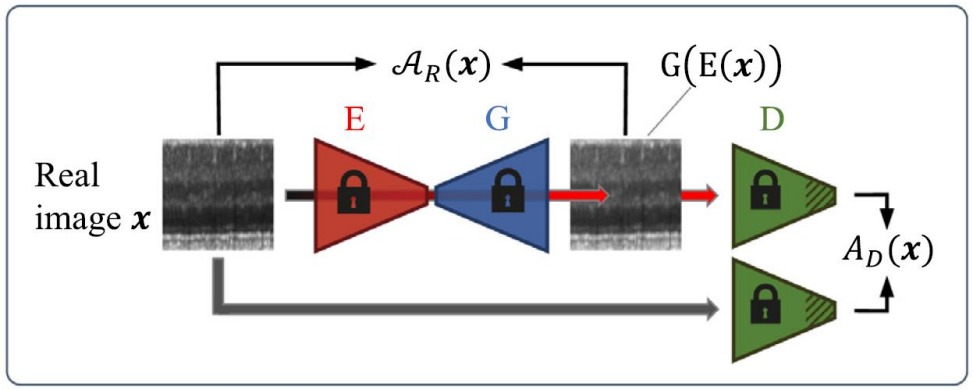
$$\mathcal{L}_{izi_f}(\mathbf{x}) = \frac{1}{n} \cdot \|\mathbf{x} - G\big(E(\mathbf{x})\big)\|^2 + \frac{\kappa}{n_d} \cdot \|f(\mathbf{x}) - f\Big(G\big(E(\mathbf{x})\big)\Big)\|^2$$

Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.

Reconstruction-based methods

- Detection of anomalies

For izi_f, the Anomaly Score (AS) comprises a discriminator feature residual error and an image reconstruction error.



$$\mathcal{A}(\mathbf{x}) = \mathcal{A}_R(\mathbf{x}) + \kappa \cdot \mathcal{A}_D(\mathbf{x})$$

where $\mathcal{A}_R(\mathbf{x}) = \frac{1}{n} \cdot \|\mathbf{x} - G\big(E(\mathbf{x})\big)\|^2$

$$\mathcal{A}_D(\mathbf{x}) = \frac{1}{n_d} \cdot \|f(\mathbf{x}) - f\Big(G\big(E(\mathbf{x})\big)\Big)\|^2$$

For izi and ziz, the AS only comprises $\mathcal{A}_R(\mathbf{x})$

Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.

Reconstruction-based methods

- Experiments on optical coherence tomography (OCT)

Comparison of investigated encoder training architectures: *ziz*, *izi* and izi_f (*f-AnoGAN*) based on the same WGAN training.

	Precision	Sensitivity	Specificity	f-score	AUC
ziz	0.7047	**0.8146**	0.8522	0.7557	0.9066
izi	0.7018	0.7497	0.8621	0.7250	0.8874
izi_f	**0.7863**	0.8091	**0.9049**	**0.7975**	**0.9301**

Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.

Reconstruction-based methods

- Experiments on optical coherence tomography (OCT)

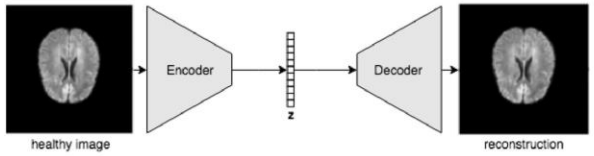
The image-level anomaly detection performance of a convolutional autoencoder (AE), adversarial convolutional autoencoder (AdvAE), ALI model, based on the output of the WGAN discriminator (A_D), iterative z-mapping utilizing the trained WGAN model (iterative), and f-AnoGAN.

	Precision	Sensitivity	Specificity	f-score	AUC
AE	0.6824	0.7195	0.8550	0.7005	0.8688
AdvAE	0.6405	0.7856	0.8092	0.7057	0.8649
ALI	0.5063	0.7434	0.6863	0.6023	0.7897
A_D	0.4909	0.6831	0.6931	0.5713	0.7504
iterative	0.7202	0.8049	0.8645	0.7602	0.9114
f-AnoGAN	**0.7863**	**0.8091**	**0.9049**	**0.7975**	**0.9301**

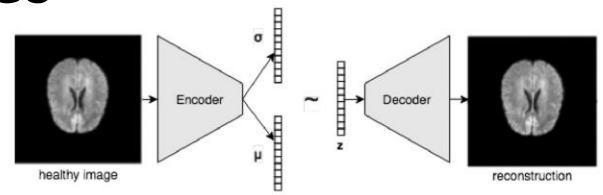
Schlegl, Thomas, et al. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. MIA 2019.

Reconstruction-based methods

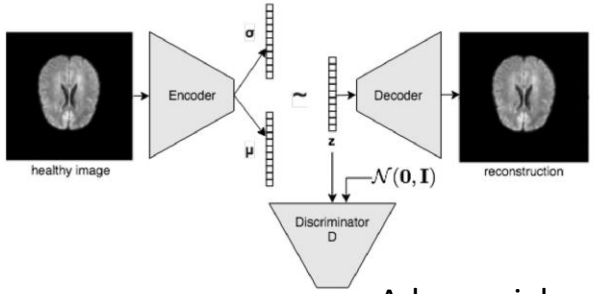
- Other reconstruction-based structures



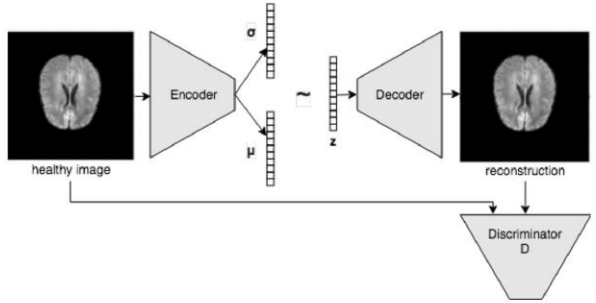
(a) AE Autoencoder [1]



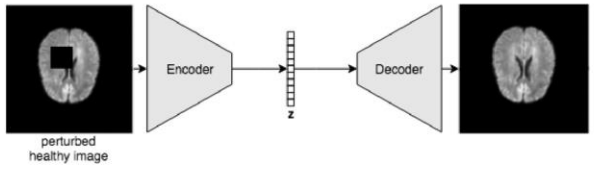
(b) VAE Variational autoencoder



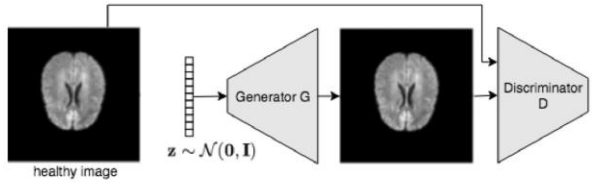
(c) AAE Adversarial autoencoder



(d) AnoVAEGAN [2]



(e) Context AE [3]



(f) GAN [4]

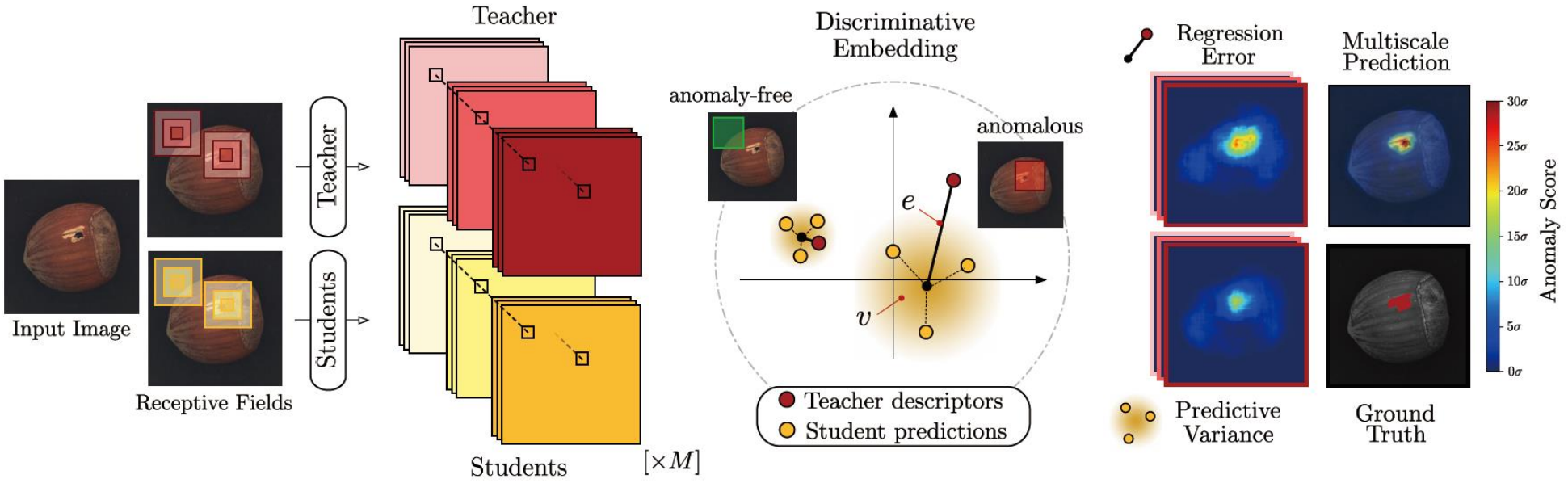
[1] Baur, et al. Autoencoders for unsupervised anomaly segmentation in brain MR images: a comparative study. MIA 2021.
[2] Baur, et al. Deep autoencoding models for unsupervised anomaly segmentation in brain MR images. MICCAI brainlesion workshop 2018.
[3] Zimmerer, et al. Context-encoding variational autoencoder for unsupervised anomaly detection. arXiv:1812.05941.
[4] Schlegl, et al. Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. IPMI 2017.

Anomaly Detection in MIA

- Introduction
- Reconstruction-based methods
- Ensemble-based methods
- Self-supervised methods
- Non-OCC settings
- Challenge and future direction

Ensemble-based methods

- Detecting anomalies based on feature discrepancies.

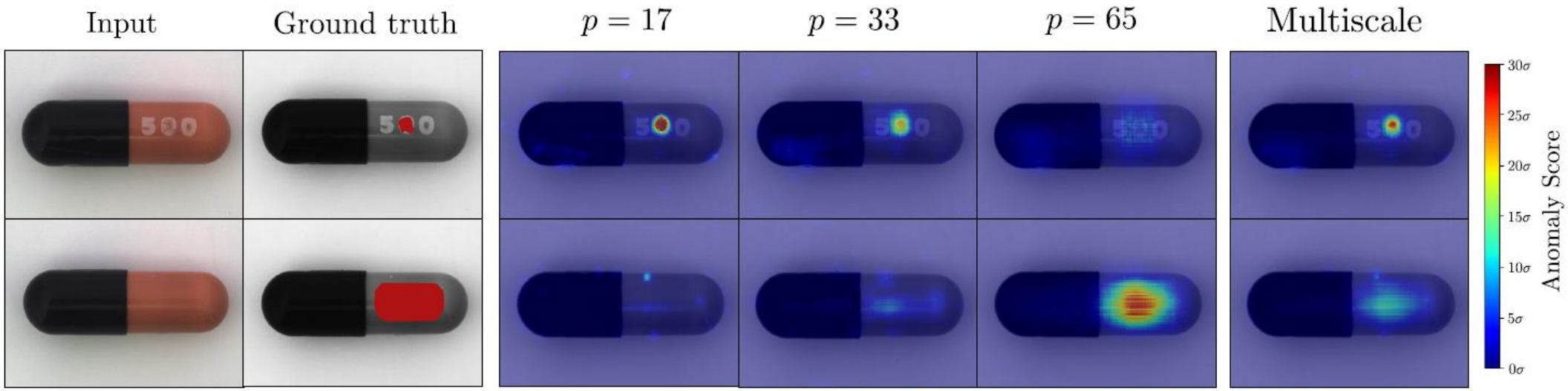


Bergmann et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. CVPR 2020.

Ensemble-based methods

- Experiments on MVTec AD dataset

Anomaly detection at multiple scales.



Bergmann et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. CVPR 2020.

Ensemble-based methods

- Performance for different receptive field sizes p.

	Category	$p = 17$	$p = 33$	$p = 65$	Multiscale
Textures	Carpet	0.795	**0.893**	0.695	0.879
	Grid	0.920	0.949	0.819	**0.952**
	Leather	0.935	**0.956**	0.819	0.945
	Tile	0.936	**0.950**	0.912	0.946
	Wood	**0.943**	0.929	0.725	0.911
Objects	Bottle	0.814	0.890	0.918	**0.931**
	Cable	0.671	0.764	**0.865**	0.818
	Capsule	0.935	0.963	0.916	**0.968**
	Hazelnut	0.971	**0.965**	0.937	**0.965**
	Metal nut	0.891	0.928	0.895	**0.942**
	Pill	0.931	0.959	0.935	**0.961**
	Screw	0.915	0.937	0.928	**0.942**
	Toothbrush	**0.946**	0.944	0.863	0.933
	Transistor	0.540	0.611	**0.701**	0.666
	Zipper	0.848	0.942	0.933	**0.951**
	Mean	0.866	0.900	0.857	**0.914**

Bergmann et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. CVPR 2020.

Ensemble-based methods

- Comparison with others

	Category	Ours $p = 65$	1-NN	OC-SVM	K-Means	ℓ_2-AE	VAE	SSIM-AE	AnoGAN	CNN-Feature Dictionary
Textures	Carpet	**0.695**	0.512	0.355	0.253	0.456	0.501	0.647	0.204	0.469
	Grid	0.819	0.228	0.125	0.107	0.582	0.224	**0.849**	0.226	0.183
	Leather	**0.819**	0.446	0.306	0.308	**0.819**	0.635	0.561	0.378	0.641
	Tile	**0.912**	0.822	0.722	0.779	0.897	0.870	0.175	0.177	0.797
	Wood	0.725	0.502	0.336	0.411	**0.727**	0.628	0.605	0.386	0.621
Objects	Bottle	**0.918**	0.898	0.850	0.495	0.910	0.897	0.834	0.620	0.742
	Cable	**0.865**	0.806	0.431	0.513	0.825	0.654	0.478	0.383	0.558
	Capsule	**0.916**	0.631	0.554	0.387	0.862	0.526	0.860	0.306	0.306
	Hazelnut	**0.937**	0.861	0.616	0.698	0.917	0.878	0.916	0.698	0.844
	Metal nut	**0.895**	0.705	0.319	0.351	0.830	0.576	0.603	0.320	0.358
	Pill	**0.935**	0.725	0.544	0.514	0.893	0.769	0.830	0.776	0.460
	Screw	**0.928**	0.604	0.644	0.550	0.754	0.559	0.887	0.466	0.277
	Toothbrush	**0.863**	0.675	0.538	0.337	0.822	0.693	0.784	0.749	0.151
	Transistor	0.701	0.680	0.496	0.399	**0.728**	0.626	0.725	0.549	0.628
	Zipper	**0.933**	0.512	0.355	0.253	0.839	0.549	0.665	0.467	0.703
	Mean	**0.857**	0.640	0.479	0.423	0.790	0.639	0.694	0.443	0.515

Bergmann et al. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. CVPR 2020.

Anomaly Detection in MIA

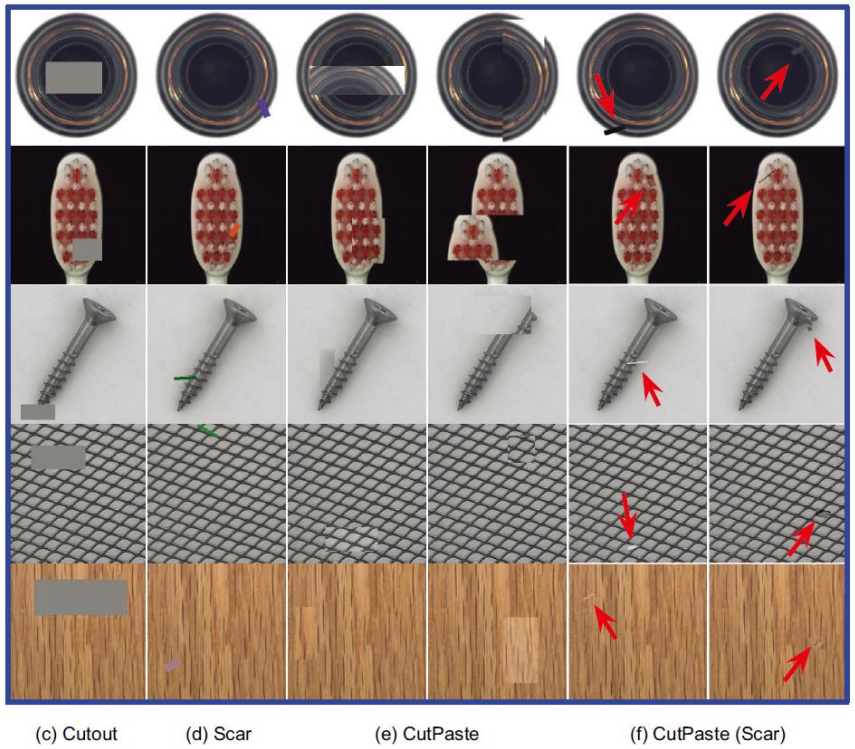
- Introduction
- Reconstruction-based methods
- Ensemble-based methods
- Self-supervised methods
- Non-OCC settings
- Challenge and future direction

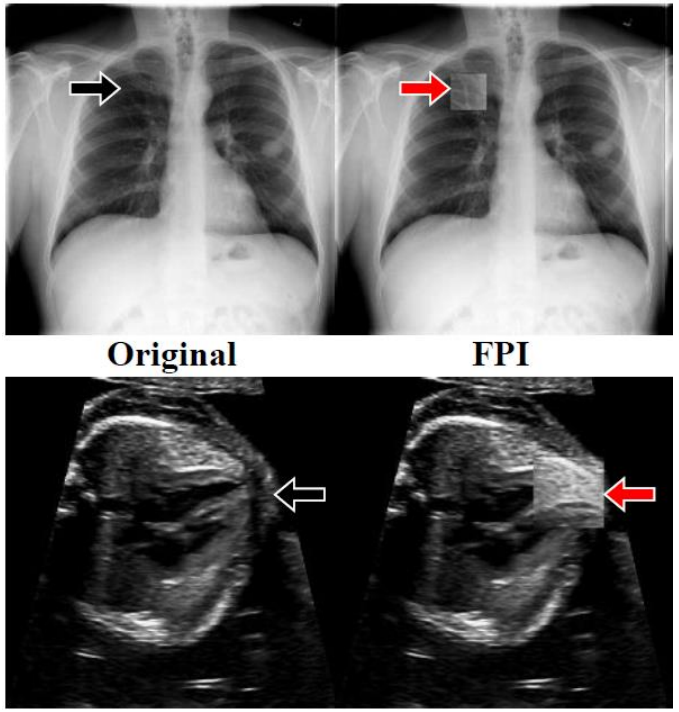
Self-supervised methods

- Self-supervised methods aim to learn more relevant representations by training on proxy tasks.

Self-supervised methods

- For anomaly detection, we can synthesize defects manually and train the network on pseudo labels.





FPI: Foreign Patch Interpolation

Li, Chun-Liang, et al. Cutpaste: Self-supervised learning for anomaly detection and localization. CVPR 2021.
Tan, Jeremy, et al. Detecting outliers with foreign patch interpolation. Journal of Machine Learning for Biomedical Imaging.

Self-supervised methods

- In order to reduce the overfitting, methods for synthesizing more “real” and subtle defects are required.

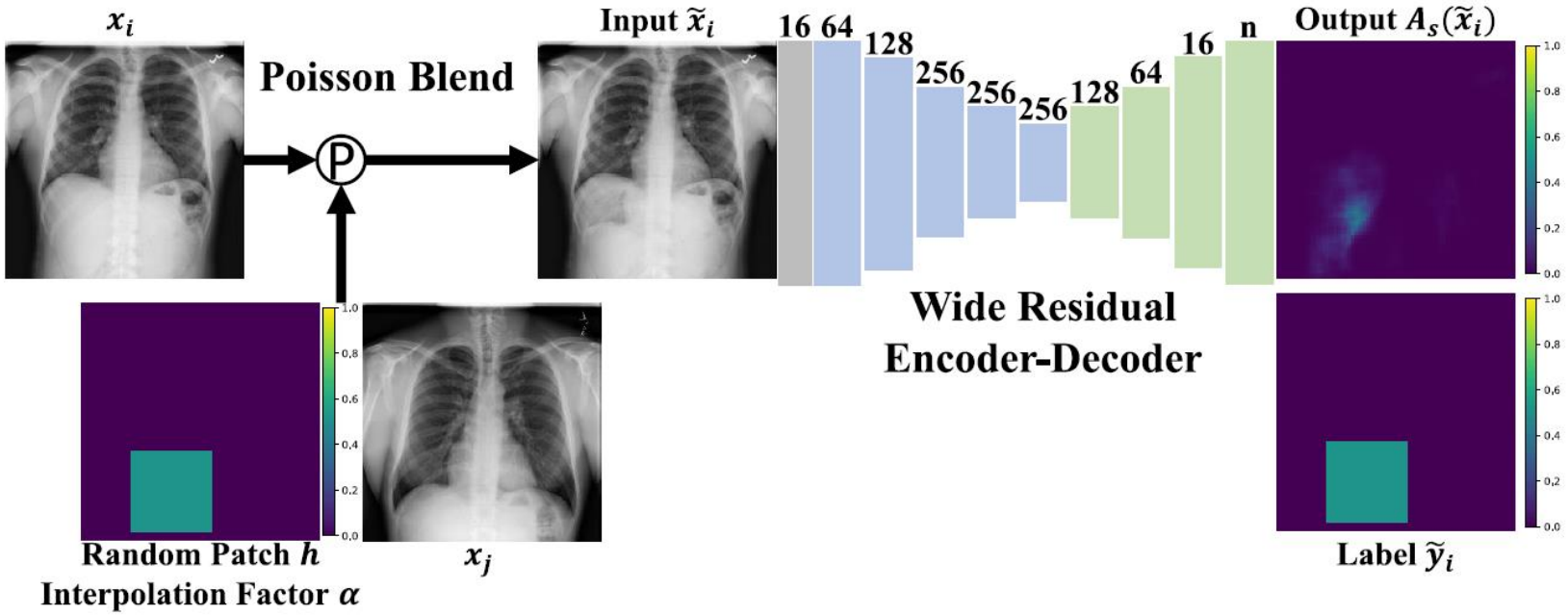
- Poisson Image Interpolation (PII)

Rather than taking the raw intensity values from the source, PII extracts the *image gradient* across the image.

Tan, Jeremy, et al. Detecting outliers with poisson image interpolation. MICCAI 2021.

Self-supervised methods

- Self-supervised training



$$\mathcal{L}_{\text{bce}} = -\tilde{y}_{i_p} log A_s(\tilde{x}_{i_p}) - (1 - \tilde{y}_{i_p}) \log(1 - A_s(\tilde{x}_{i_p}))$$

Tan, Jeremy, et al. Detecting outliers with poisson image interpolation. MICCAI 2021.

Self-supervised methods

- For PII, blending the content of a source image (x_j) into the context of a destination image (x_i), the goal is to find f_{in} for the Poisson Equation with Dirichlet boundary conditions (at the edge of the patch):

$$\min_{f_{in}} \iint_{h} |\nabla f_{in} - \mathbf{v}|^2 \text{ with } f_{in}\Big|_{\partial h} = f_{out}\Big|_{\partial h}$$

f_{in}: intensity values within the patch h.

f_{out}: intensity values of destination image outside h.

$\mathbf{v}$: the gradient of source image.

Tan, Jeremy, et al. Detecting outliers with poisson image interpolation. MICCAI 2021.

Self-supervised methods

- Understanding:

$$\min_{f_{in}} \iint_{h} |\nabla f_{in} - \mathbf{v}|^2 \text{ with } f_{in}\Big|_{\partial h} = f_{out}\Big|_{\partial h}$$

f_{in} should:

1. match the surrounding values f_{out} of the destination image, along the border of the patch h.
2. follow the relative changes (image gradient), **v**, of the source image.

Tan, Jeremy, et al. Detecting outliers with poisson image interpolation. MICCAI 2021.

Self-supervised methods

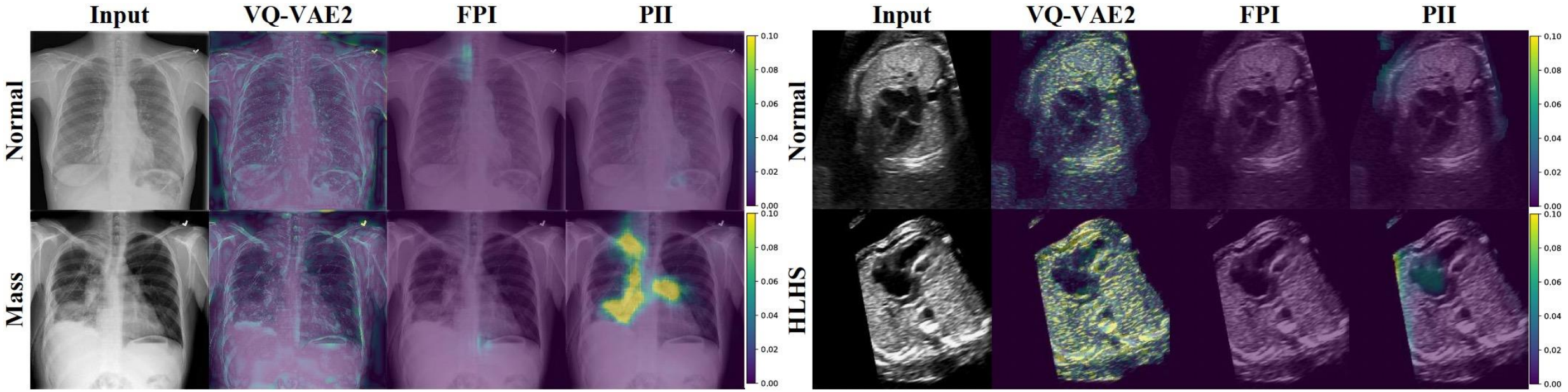
- Experiments on Chest X-ray and Fetal US Dataset.

Dataset	Chest X-ray		Fetal US	
	♂PA	**♀PA**	**4CH**	**3VT**
	Number of Images			
Normal Train	17852	14720	283×20	225×20
Normal Test	2634	2002	34×20	35×20
Anomalous Test	3366	2748	54×20	38×20
	Average Precision			
Deep SVDD	0.565	0.556	0.685	0.893
VQ-VAE2	0.503	0.516	0.617	0.578
FPI	0.533	0.586	0.658	0.710
PII	**0.690**	**0.703**	**0.723**	**0.929**

Tan, Jeremy, et al. Detecting outliers with poisson image interpolation. MICCAI 2021.

Self-supervised methods

- Experiments on Chest X-ray and Fetal US Dataset.



Examples of chest X-ray (left) and ultrasound (right) images with pixelwise anomaly scores from each method.

Tan, Jeremy, et al. Detecting outliers with poisson image interpolation. MICCAI 2021.

Anomaly Detection in MIA

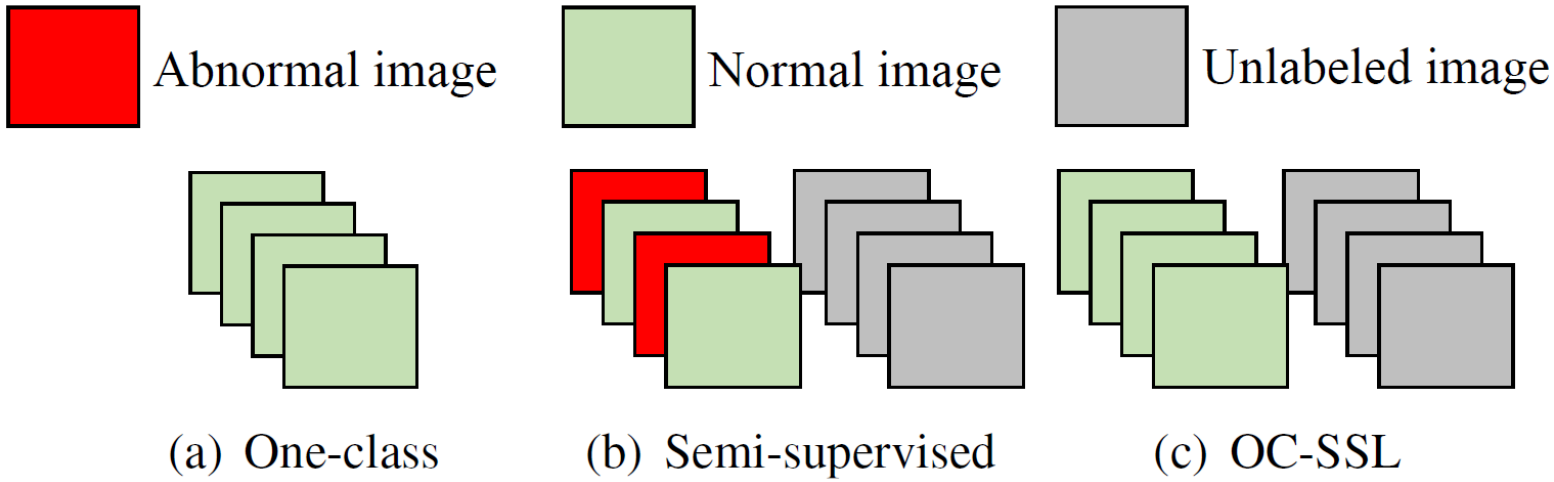
- Introduction
- Reconstruction-based methods
- Self-supervised methods
- Ensemble-based methods
- Non-OCC settings
- Challenge and future direction

Non-OCC settings

- The One-class Classification (OCC) setting doesn’t always fit the medical scenarios well.
- In addition to normal images, plenty of unlabelled data (comprising both normal and abnormal samples) are readily available in clinical practice, which are not exploited by methods under the OCC setting.

Non-OCC settings

- To exploit the unlabelled medical images for anomaly detection, one-class semi-supervised learning (OC-SSL) is proposed.

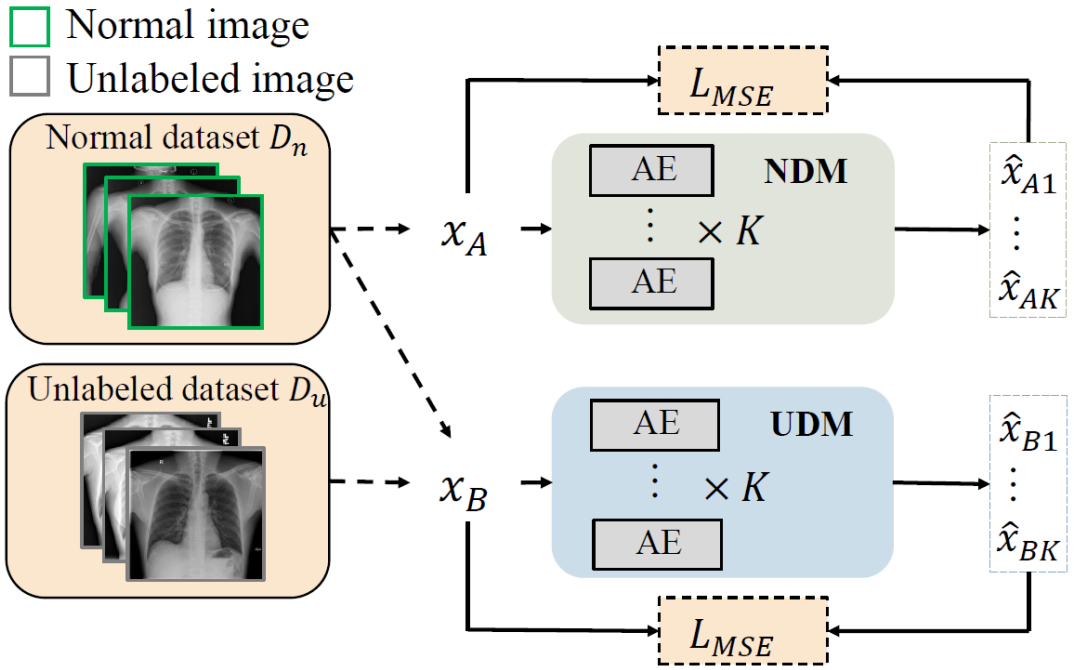


- Train a model on a normal dataset D_n and an unlabelled dataset D_u.

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Model the distribution of normal and unlabelled training data using ensembles of reconstruction networks.



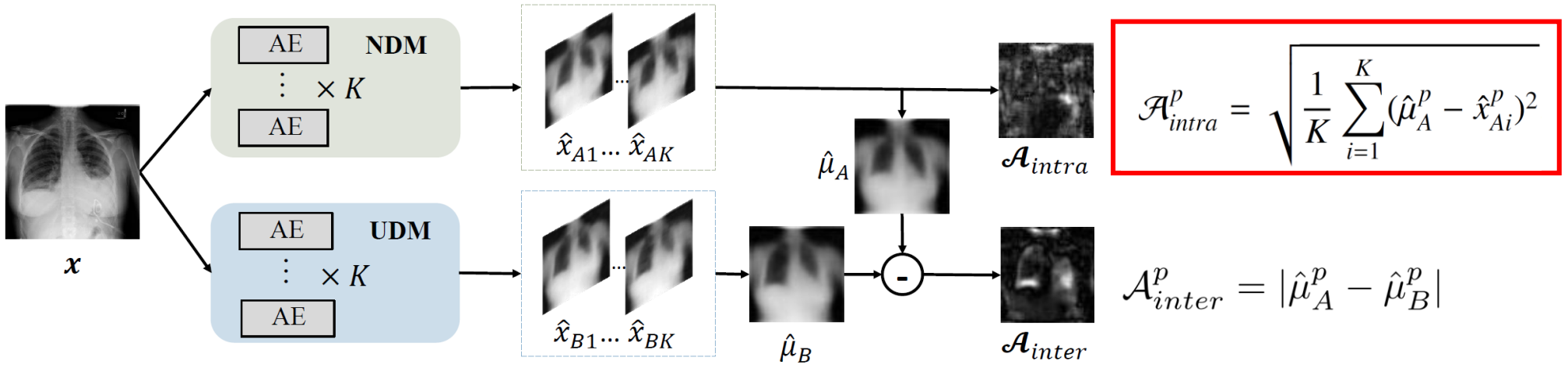
NDM: normative distribution module

UDM: unknown distribution module

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Inference: discrepancy among reconstructions of ensemble networks are utilized as anomaly score.

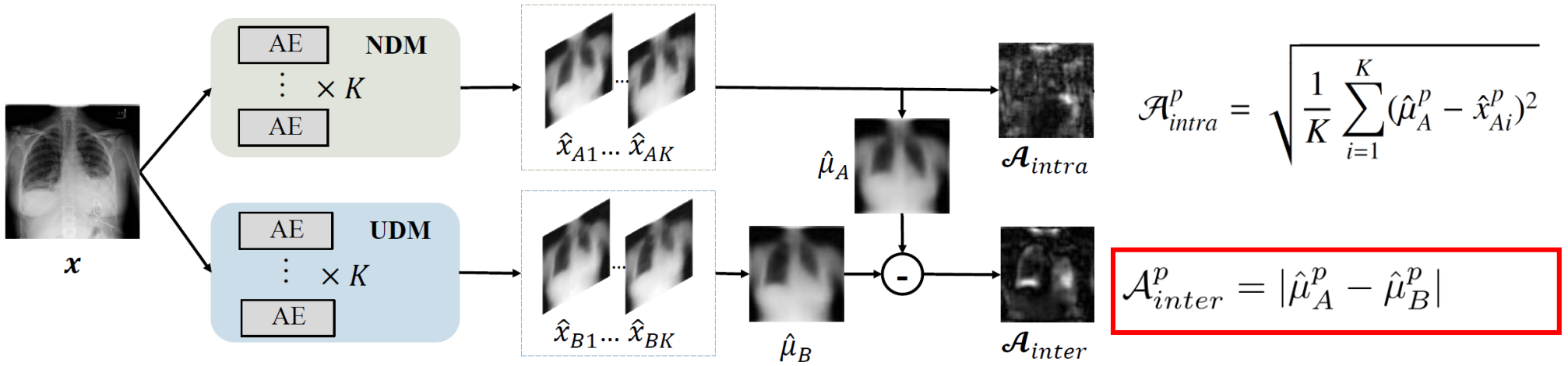


As NDM never sees abnormal images, it will express high intra-discrepancy on unseen abnormal regions.

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Inference: discrepancy among reconstructions of ensemble networks are utilized as anomaly score.

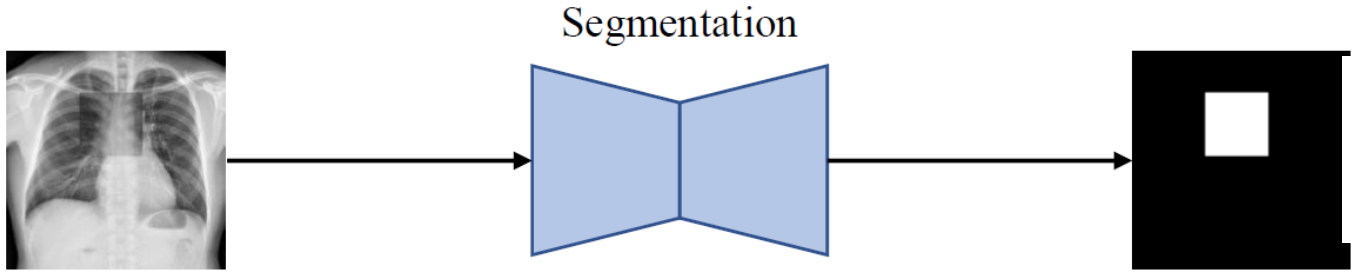


As UDM captures some anomalous information from unlabeled images, it will perform differently with NDM in abnormal regions

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

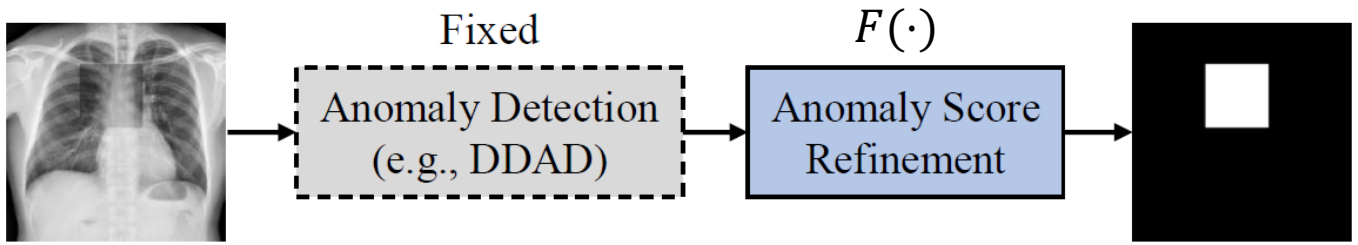
Non-OCC settings

- Self-supervised anomaly score refinement net is designed to further refine the predicted score maps.



$$\mathcal{A}^p_{intra} = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(\hat{\mu}^p_A - \hat{x}^p_{Ai})^2}$$

(a) Standard self-supervised anomaly detection



$$\mathcal{R}_{intra} = F(\mathcal{A}_{intra})$$

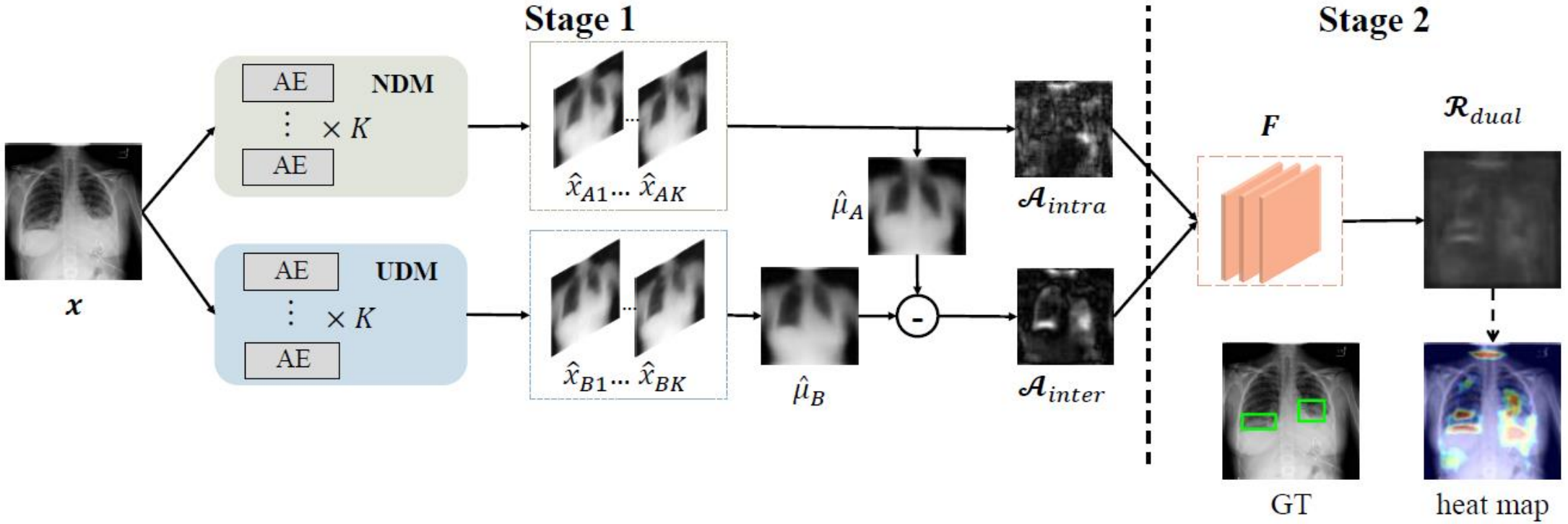
$$\mathcal{R}_{dual} = F([\mathcal{A}_{intra}, \mathcal{A}_{inter}])$$

(b) The proposed self-supervised anomaly score refinement

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Self-supervised anomaly score refinement net is designed to further refine the predicted score maps.



[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Medical Anomaly Detection Benchmark

Table 1. Summary of dataset repartitions. Note that D_u is built using data selected from the images presented in parentheses without the use of their annotations.

Dataset	Repartition		
	Normal Dataset D_n	Unlabeled Dataset D_u	Testing Dataset D_t
RSNA[1]	3851	4000 (4000 normal + 5012 abnormal images)	1000 normal + 1000 abnormal images
VinDr-CXR[2] (Nguyen et al., 2022)	4000	4000 (5606 normal + 3394 abnormal images)	1000 normal + 1000 abnormal images
CXAD	2000	2000 (800 normal + 1200 abnormal images)	499 normal + 501 abnormal images
Brain Tumor[3]	1000	1000 (400 normal + 2666 abnormal images)	600 normal + 600 abnormal images
LAG (Li et al., 2019)	1500	1500 (832 normal + 900 abnormal images)	811 normal + 811 abnormal images

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

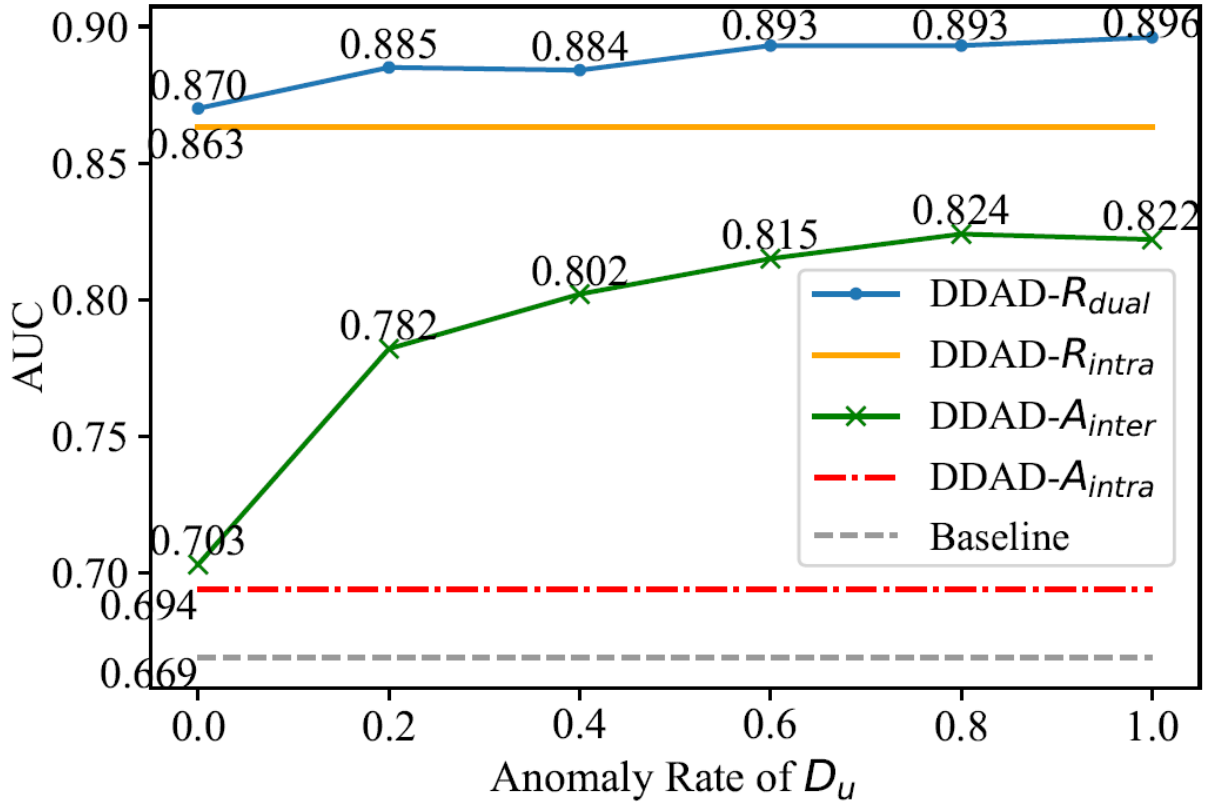
- Comparison with state-of-the-arts

Unlabeled data	Method	Taxonomy	RSNA		VinDr-CXR		CXAD		Brain MRI		LAG	
			AUC	AP	AUC	AP	AUC	AP	AUC	AP	AUC	AP
✗	AE	Rec.	66.9	66.1	55.9	60.3	55.6	59.6	79.7	71.9	79.3	76.1
	MemAE (Gong et al., 2019)	Rec.	68.0	67.1	55.8	59.8	56.0	60.0	77.4	70.0	78.5	74.9
	Ganomaly (Akcay et al., 2018)	Rec.	71.4	69.1	59.6	60.3	62.5	63.0	75.1	69.7	77.7	75.7
	DRAEM (Zavrtanik et al., 2021)	Rec.+Self-sup.	62.3	61.6	63.0	68.3	54.3	55.6	72.1	64.6	47.2	49.0
	CutPaste$^{\text{IN-Pretr.}}$ (Li et al., 2021)	Self-sup.+GDE	79.4	74.4	70.2	69.8	53.6	57.3	92.0	89.4	69.1	64.6
	CutPaste$^{\text{Scrat.}}$ (Li et al., 2021)	Self-sup.+GDE	75.1	72.6	59.6	58.6	50.3	53.6	92.0	89.9	63.4	59.8
	CutPaste (e2e) (Schlüter et al., 2022)	Self-sup.	55.0	58.0	54.6	55.5	47.0	48.4	71.0	66.8	53.7	53.9
	FPI (Tan et al., 2020)	Self-sup.	47.6	55.7	48.2	49.9	44.8	47.6	83.1	78.9	53.4	55.6
	PII (Tan et al., 2021)	Self-sup.	82.9	83.6	65.9	65.8	52.7	53.7	84.3	80.5	61.0	60.7
	NSA (Schlüter et al., 2022)	Self-sup.	82.2	82.6	64.4	65.8	58.5	58.2	84.4	81.1	67.9	67.0
	f-AnoGAN (Schlegl et al., 2019)	Rec.	79.8	75.6	76.3	**74.8**	61.9	67.3	82.5	74.3	84.2	77.5
	IGD (Chen et al., 2022)	Rec.	81.2	78.0	59.2	58.7	55.2	57.6	**94.3**	90.6	80.7	75.3
	AE-U (Mao et al., 2020)	Rec.	86.7	84.7	73.8	72.8	66.4	66.9	94.0	89.0	81.3	78.9
	Ours (AE), $\mathcal{R}_{intra}$	Ens.+Self-sup.	86.3	85.5	77.2	74.2	63.8	65.4	85.0	77.6	79.5	74.5
	Ours (MemAE), $\mathcal{R}_{intra}$	Ens.+Self-sup.	87.2	86.1	73.9	72.1	62.4	64.5	82.9	78.6	80.1	77.6
	Ours (AE-U), $\mathcal{R}_{intra}$	Ens.+Self-sup.	**88.3**	**87.6**	**78.2**	74.6	**69.4**	**69.3**	94.2	**91.9**	**86.0**	**84.0**
✓	CutPaste (e2e)* (Schlüter et al., 2022)	Self-sup.	59.8	61.7	59.2	60.0	48.9	50.7	69.8	64.9	48.9	51.7
	FPI* (Tan et al., 2020)	Self-sup.	46.6	53.8	47.4	49.4	45.3	47.6	86.6	83.8	52.9	56.1
	PII* (Tan et al., 2021)	Self-sup.	84.3	85.4	66.8	67.2	54.4	54.8	90.0	89.1	63.1	63.1
	NSA* (Schlüter et al., 2022)	Self-sup.	84.2	84.3	64.4	64.8	57.4	57.0	88.8	84.7	68.6	68.0
	Ours (AE), $\mathcal{R}_{dual}$	Ens.+Self-sup.	89.3	89.5	77.4	77.7	65.0	67.2	93.0	87.1	89.0	86.9
	Ours (MemAE), $\mathcal{R}_{dual}$	Ens.+Self-sup.	88.5	87.8	75.3	74.1	63.5	64.3	91.4	84.8	88.7	86.5
	Ours (AE-U), $\mathcal{R}_{dual}$	Ens.+Self-sup.	**91.3**	**91.6**	**85.9**	**84.3**	**71.0**	**72.7**	**97.2**	**95.2**	**93.1**	**92.3**

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Ablation study: performance with different anomaly rates



[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Ablation study: performance on seen and unseen pathologies.

Table 4. Performance of DDAD on seen and unseen pathologies. Setting A indicates the testing set contains only pathologies in $\mathcal{P}_A$, which could appear in D_u. Setting B indicates the testing set contains only pathologies in $\mathcal{P}_B$, which are unseen in D_u.

Method	Unlabeled dataset D_u	Setting A		Setting B	
		AUC (%)	AP (%)	AUC (%)	AP (%)
Reconstruction	0	49.7	55.6	63.7	70.1
DDAD-$\mathcal{A}_{inter}$	4000 (4000 normal + 0 abnormal images)	54.0	60.0	66.0	71.1
	4000 (2412 normal + 1588 abnormal images in $\mathcal{P}_A$)	$64.2^{+10.2}$	$70.8^{+10.8}$	$70.0^{+4.0}$	$75.8^{+4.7}$

[1] Cai, et al. Dual-Distribution Discrepancy for Anomaly Detection in Chest X-Rays. MICCAI 2022.
[2] Cai, et al. Dual-distribution discrepancy with self-supervised refinement for anomaly detection in medical images. MIA 2023.

Non-OCC settings

- Visualization

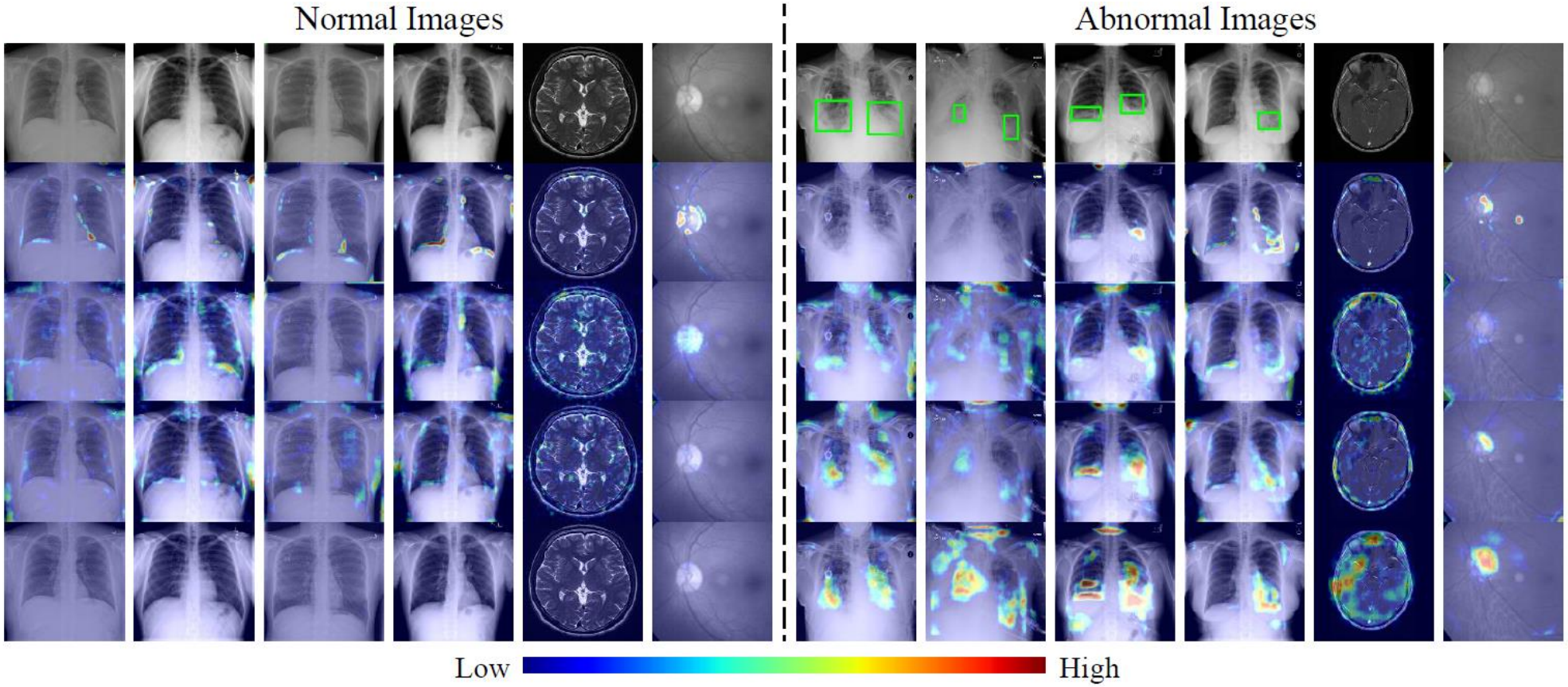


Fig. 9. Visualization of heat maps on medical datasets. From top to bottom: Original images, heat maps of $\mathcal{A}_{rec}$, heat maps of $\mathcal{A}_{intra}$, heat maps of $\mathcal{A}_{inter}$, heat maps of $\mathcal{R}_{dual}$. The green bounding boxes indicate abnormal regions.

Anomaly Detection in MIA

- Introduction
- Reconstruction-based methods
- Self-supervised methods
- Ensemble-based methods
- Non-OCC settings
- Challenge and future direction

Challenge and future directions

- The Necessity of Benchmark Datasets

1. It is debatable whether such methods utilizing only normal datasets can be called unsupervised or should be seen as weakly-supervised?
2. The community should aim for methods trained from all kinds of samples, even data potentially including anomalies, without the need for human ratings.
3. Many studies evaluated their methods on datasets with different settings (e.g., lesion types) using various evaluation metrics, which makes it hard to compare them directly and fairly.

Challenge and future directions

- Lack of Generalization

In medical domain, there are no large-scale datasets such as ImageNet. Most datasets are also highly curated, collected in controlled environments or restricted settings that do not capture the real data distribution.

- Lack of Interpretability

Increase the transparency of the anomaly detection mechanism could help illustrate the reason behind its prediction.

- Real-world Deployment

Anomaly detection has been widely used in industrial scenarios while needs to be further explored in the medical case, such as rare disease, regular screening, etc.