Reconstruction-free Anomaly Detection with Attention-based Diffusion (RADAR)

Mehrdad Moradi[1]

Kamran Paynabar[1], Marco Grasso[2], Bianca Maria Colosimo[2,]

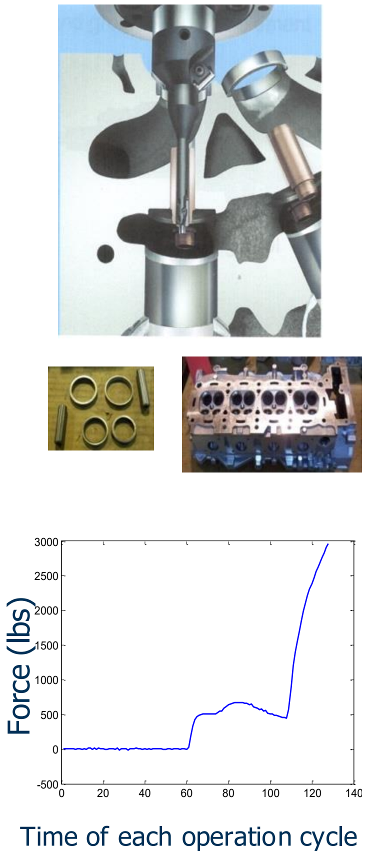
1. H. Milton Stewart School of Industrial and Systems Eng. And ML center, Georgia Institute of Technology
2. Mechanical Engineering Department, Polytechnic University of Milan

2025 INFORMS Annual Meeting, Atlanta, Georgia
October 2025

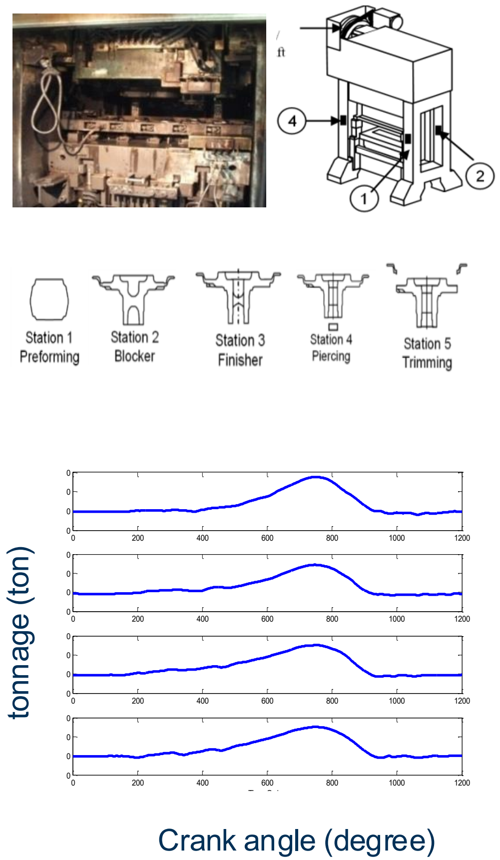
Georgia Tech

Motivation

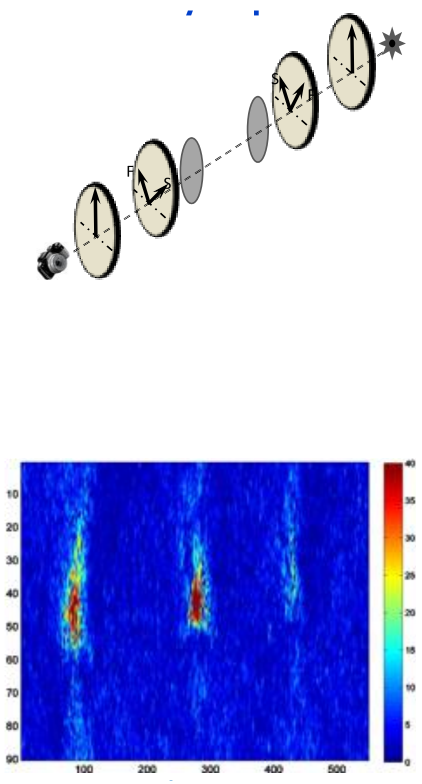
Valve Seat Assembly



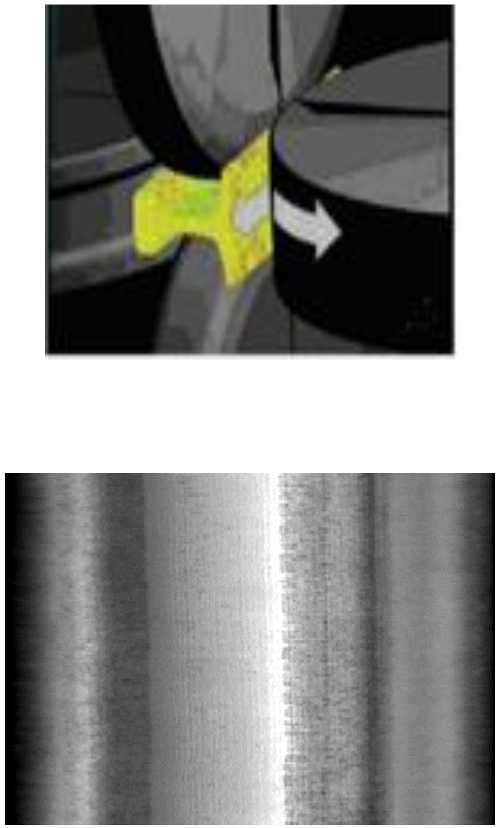
Forging



Semiconductor



Rolling

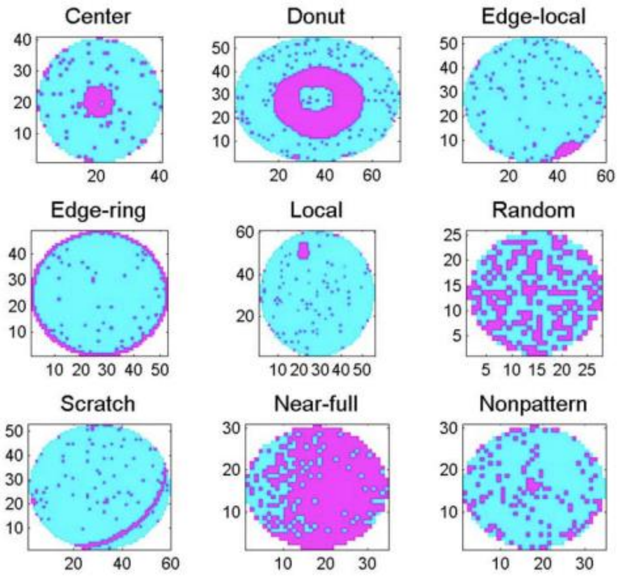


Anomaly Detection in HD Data has wide applications in different domains.

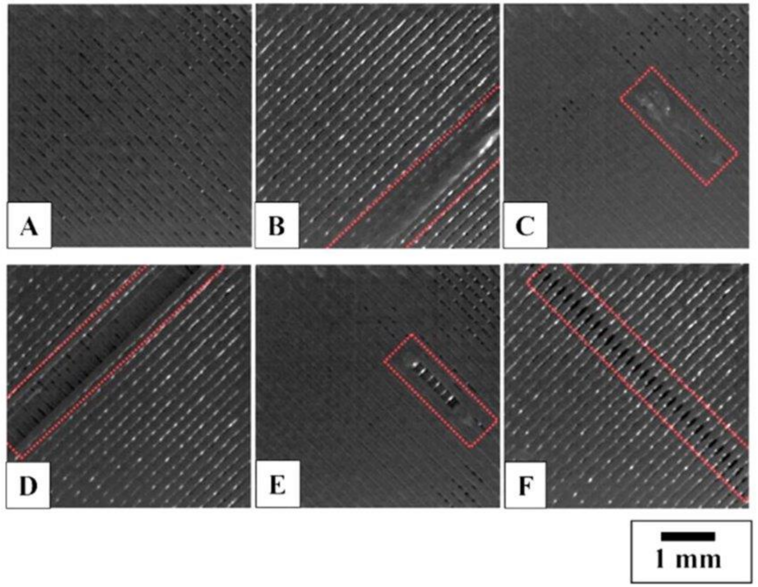
Common Methods for Anomaly Detection in HDD

- SPC-based Methods (ST-SSD and Stochastic Textured Surface)
  - Do not require large training data, but they cannot deal with complicated patterns.
  - Application specific and need to be modified for each application.

- Low-rank Decomposition Methods (e.g., RPCA [1] and SSD [2])
  - Do not require large training data, but they cannot deal with complicated patterns.
  - Low-rank (linear projection to LD space) or smoothness assumptions may not be valid.



Wafer failure map in semiconductor man. [11]

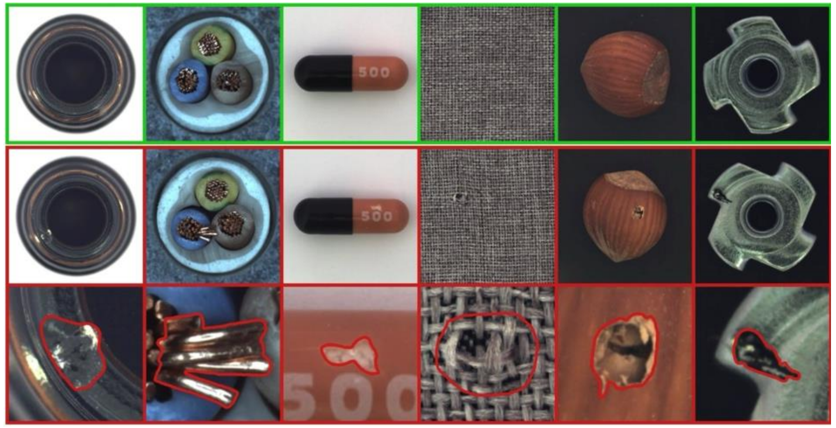


Additive manufacturing printed layers [8]

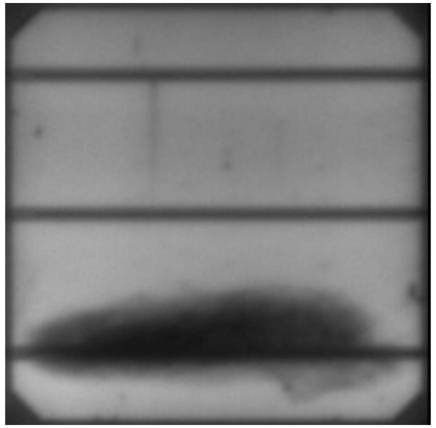
Common Methods for Anomaly Detection in HDD

- Deep Learning Approaches
  - Needs large datasets to train
  - In some cases, labeled data can be very expensive to collect

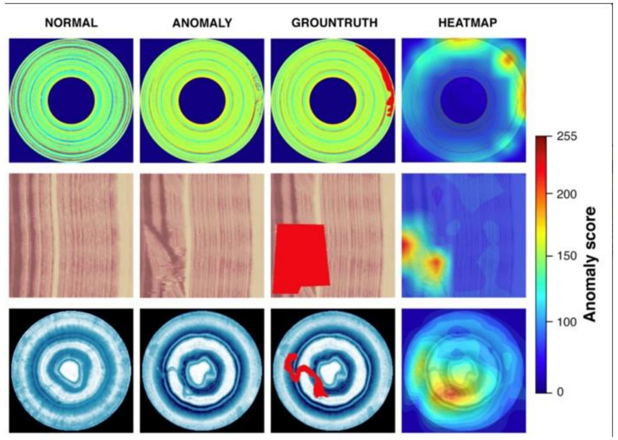
- Generative AI Approaches (e.g., VAE, GAN, DDPM)
  - Needs large datasets to train, otherwise overfits very easily
  - GANs are often unstable
  - DDPM Reconstruction-based methods cannot detect subtle defects



MVTec anomaly detection dataset [5]



ELPV data sets: solar cells
Picture of a mono crystal cell [4]

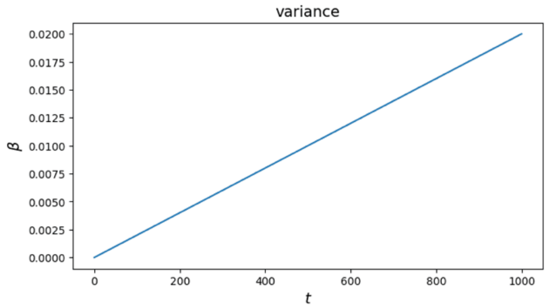


BTAD: industrial anomaly data set [10]

Review: Denoising Diffusion Probabilistic Models (DDPM [6])

- In DDPM:
  - Noise is added in a predefined Markovian chain to turn the data into pure Gaussian noise
  - Backward conditional distribution is learned as a Gaussian distribution based on MLE using neural networks.

$$p_\theta(x_{t-1} \mid x_t) = \mathcal{N}\left(x_t; \mu(x_t, t), \Sigma(x_t, t)\right)$$



$$q(x_t \mid x_{t-1}) = \mathcal{N}\left(x_t; \sqrt{1-\beta_t}x_{t-1}, \beta_t\mathbf{I}\right)$$

$q(x_t|x_{t-1})$: Forward diffusion process. β_t: predefined noise schedule.

$p_\theta(x_{t-1}|x_t)$: Backward conditional distribution. $p_\theta(x_{t-1}|x_t) = \mathcal{N}(x_{t-1}, \mu_\theta(x_t, t), \sum_\theta(x_t, t))$

$$\mu_\theta(x_t, t) = \frac{1}{\sqrt{\alpha_t}}(x_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\epsilon_\theta(x_t, t)) \Rightarrow x_{t-1} = \frac{1}{\sqrt{a_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}}}\epsilon_\theta(x_t, t)) + \sqrt{\beta_t}\epsilon$$

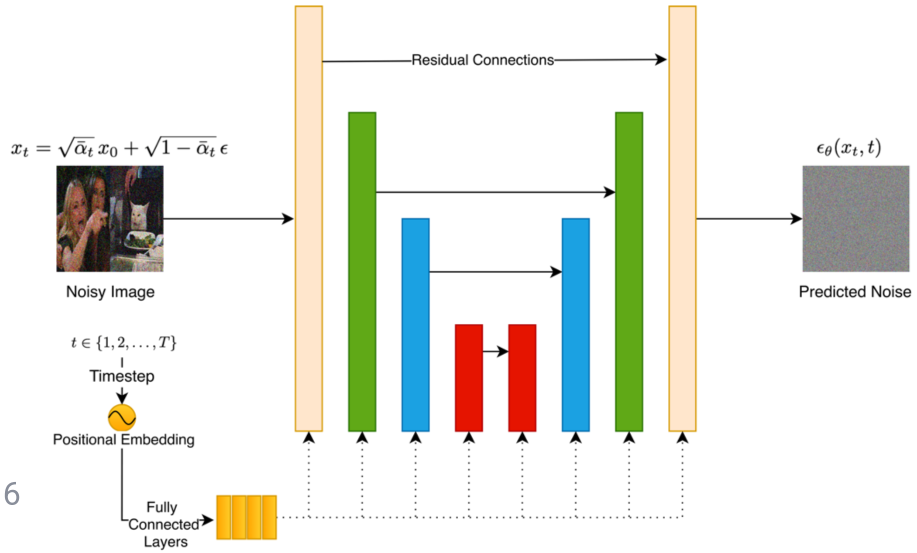
DDPM Training Algorithm [6]

- Denoising Diffusion Probabilistic Models (DDPM) using noise prediction models

$$x_{t-1} = \frac{1}{\sqrt{a_t}}(x_t - \frac{\beta_t}{\sqrt{1-\bar{\alpha}}}\epsilon_\theta(x_t, t)) + \sqrt{\beta_t}\epsilon$$

- In each training iteration:
  - Sample a data point, a time step in forward diffusion process (T is usually 1000) and a Gaussian noise and construct the noised data based on forward diffusion process
  - Apply gradient descent algorithm on the L2 norm of noise prediction error

- The architecture used is usually UNet an autoencoder with skip connections



Algorithm 1 Training

1: **repeat**
2: $\quad \mathbf{x}_0 \sim q(\mathbf{x}_0)$
3: $\quad t \sim \mathrm{Uniform}(\{1, \ldots, T\})$
4: $\quad \boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
5: $\quad$ Take gradient descent step on
$$\nabla_\theta \left\| \boldsymbol{\epsilon} - \boldsymbol{\epsilon}_\theta(\sqrt{\bar{\alpha}_t}\mathbf{x}_0 + \sqrt{1-\bar{\alpha}_t}\boldsymbol{\epsilon}, t) \right\|^2$$
6: **until** converged

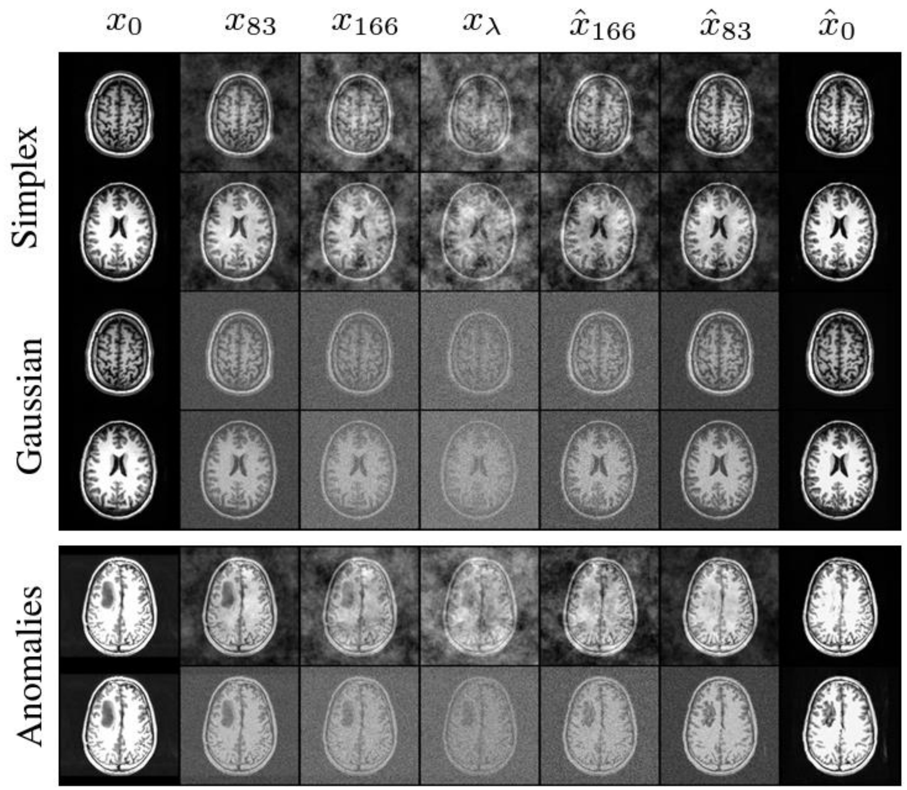
Algorithm 2 Sampling

1: $\mathbf{x}_T \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$
2: **for** $t = T, \ldots, 1$ **do**
3: $\quad \mathbf{z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ if $t > 1$, else $\mathbf{z} = \mathbf{0}$
4: $\quad \mathbf{x}_{t-1} = \frac{1}{\sqrt{\alpha_t}}\left(\mathbf{x}_t - \frac{1-\alpha_t}{\sqrt{1-\bar{\alpha}_t}}\boldsymbol{\epsilon}_\theta(\mathbf{x}_t, t)\right) + \sigma_t \mathbf{z}$
5: **end for**
6: **return** $\mathbf{x}_0$

Anomaly detection with DDPM [7]

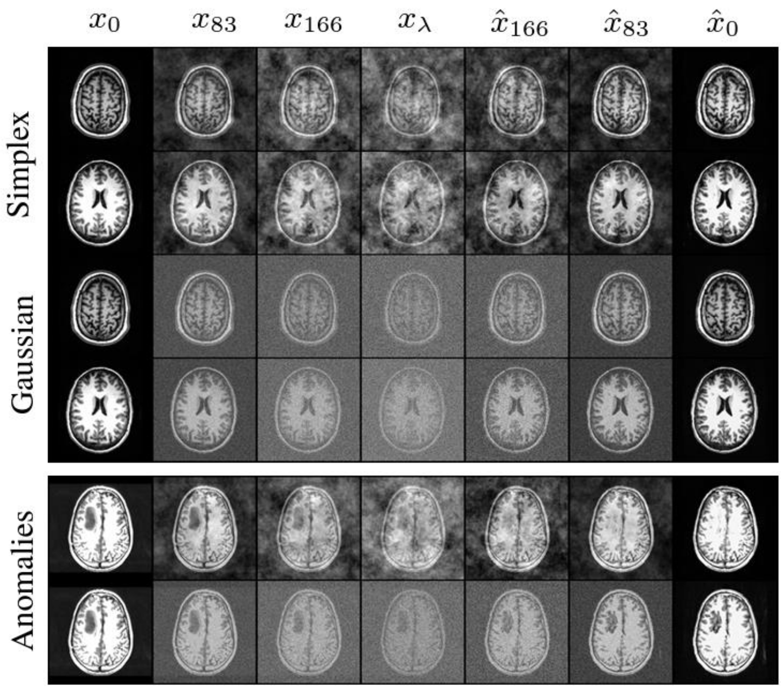
- First training DDPM on healthy data with 1000 steps in forward diffusion
- For a new image we noise image to λ steps (250 chosen)
- Apply sampling algorithm from noisy image to step 0 and get the corresponding healthy data point
- Find the difference between the original image and the reconstructed one.
- Reconstruction-based anomaly detection

- Needs large amount of healthy data
- Reconstruction-based method does not work for very subtle defects
- Cannot reconstruct well when random patterns exist like bright points



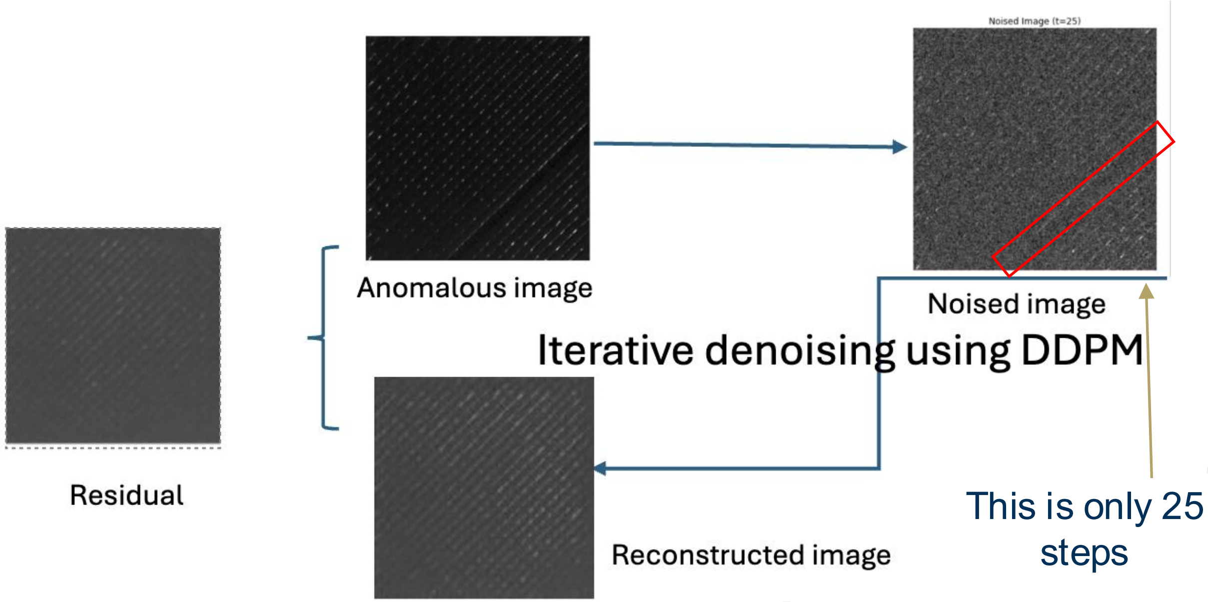
Anomaly detection with DDPM [7]

- First training DDPM on healthy data with 1000 steps in forward diffusion
- For a new image we noise image to λ steps (250 chosen)
- Apply sampling algorithm from noisy image to step 0 and get the corresponding healthy data point
- Find the residuals between the original image and the reconstructed one.
- Reconstruction-based anomaly detection



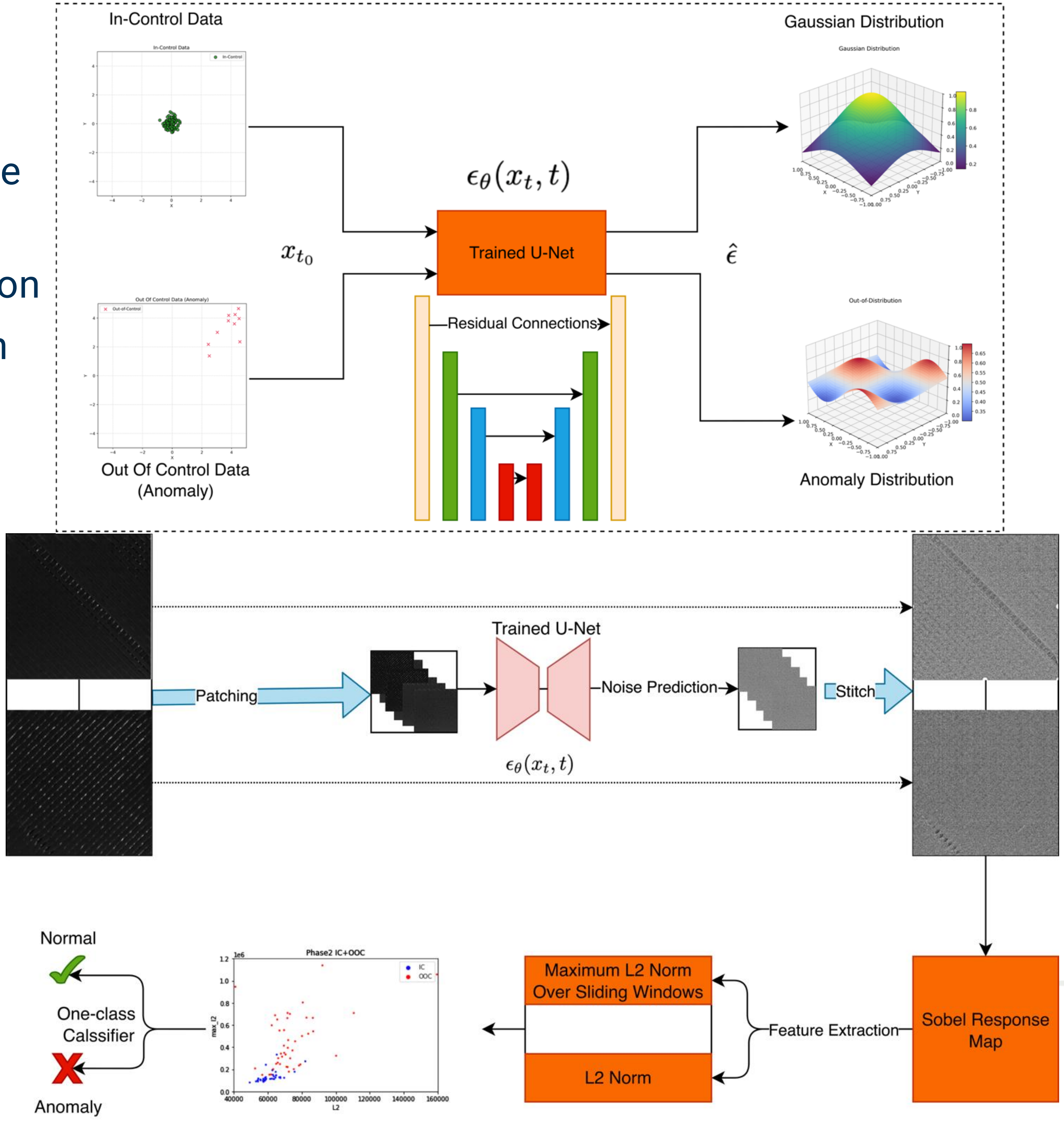
Drawbacks:

- Needs large amount of healthy data
- Reconstruction-based method does not work for subtle defects



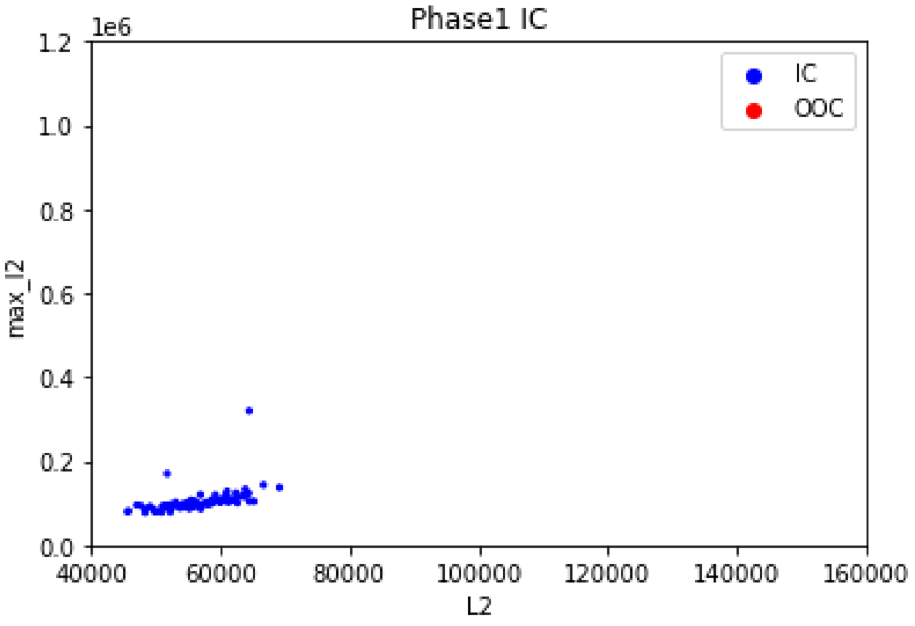
Proposed RADAR [12]

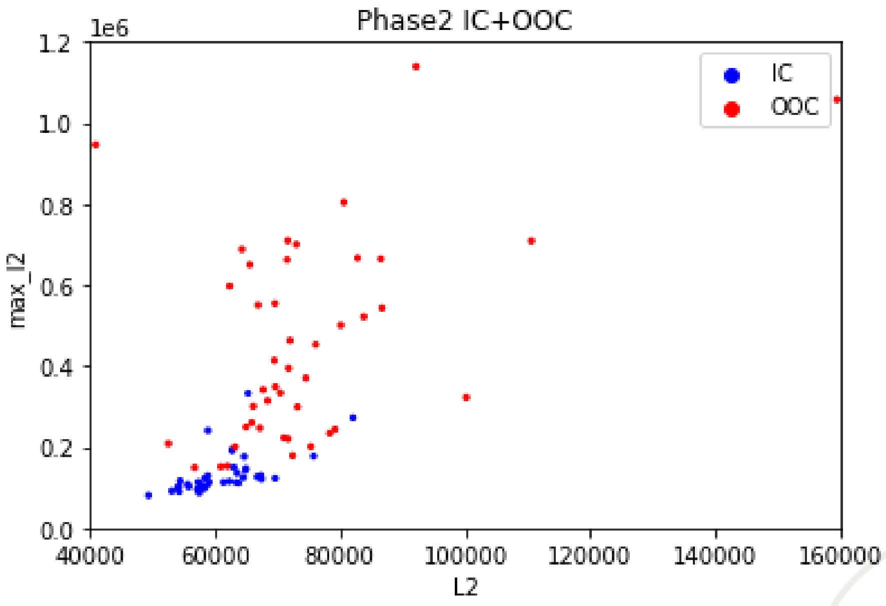
- Diffusion model is trained to predict the Gaussian noise added to the sampled image
  - In control image: approximately Gaussian prediction
  - Out of control: different distribution than Gaussian
- RADAR
  - Divide the image into patches and learn the distribution by diffusion models.
  - Significantly increases the training size
    - A 255*255 image turns into 228*228 patches of 28*28 images
    - Prevents memorization and overfitting
    - Reduces computational burden
  - In inference, noise patches in 1 step forward diffusion are predicted
  - Apply a combination of edge detection and norm-based feature extraction to extract features



Feature Extraction and One-class Classification

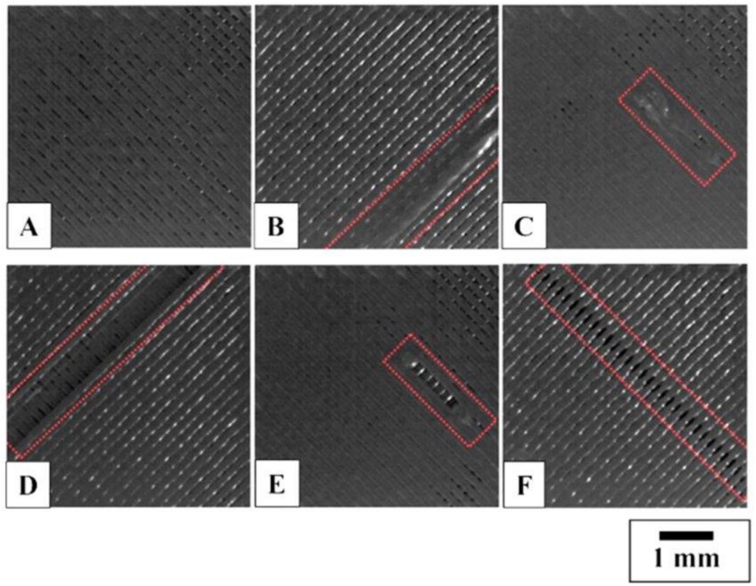
- Apply a Gaussian blur to smooth the image (kernel_size=5)
- Apply a Sobel edge detection (kernel_size=5) and extract the total L2 norm + max of L2 norm for windows with size 20 sliding 20 times in x and y axis
- For each image extract both max_L2 and L2 as two features for SPC
- Apply LOF one-class classification algorithm for anomaly detection

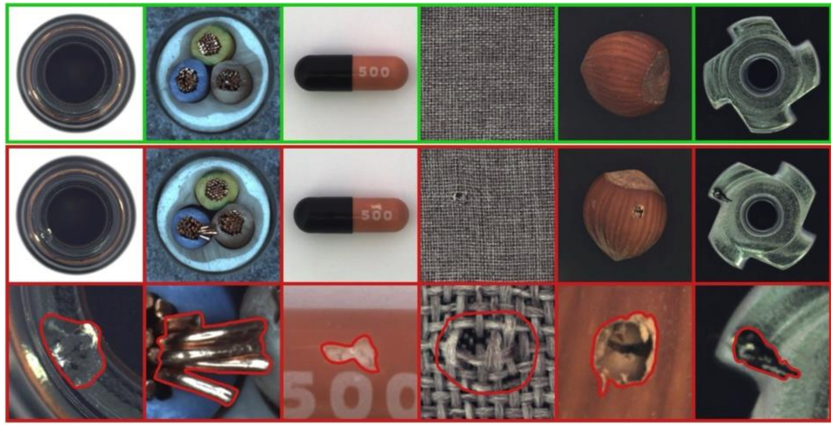




Case Study

- Dataset:
  - LPBF Additive Manufacturing Dataset (different process parameters and scan strategies)
  - MVTec-AD anomaly detection dataset: tile category
- Benchmarks:
  - State of the art diffusion models: AnoDDPM [7], DiffusionAD [9]
  - Statistical machine learning models: C&B [8], B&A [3]
  - Statistical descriptors: GLCM, Entropy, Hough Transform, SSIM
- Metrics:
  - accuracy, precision, recall, F1 score

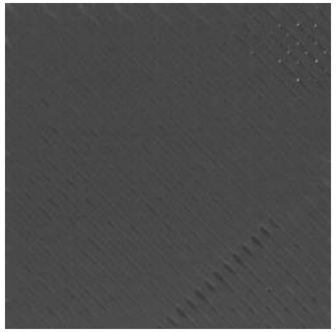




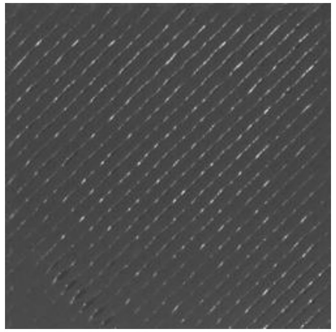
Extrusion-Based Additive Manufacturing [8]

- Phase 1 data (training): 81
  - 45 degrees orientation: 41
  - 135 degrees orientation: 40
- Phase 2 data (validation and testing): 84
  - 45 degrees in control: 21
  - 45 degrees out of control: 22
  - 135 in control: 19
  - 135 out of control: 22

In control data 45 degrees



Out of control data 135

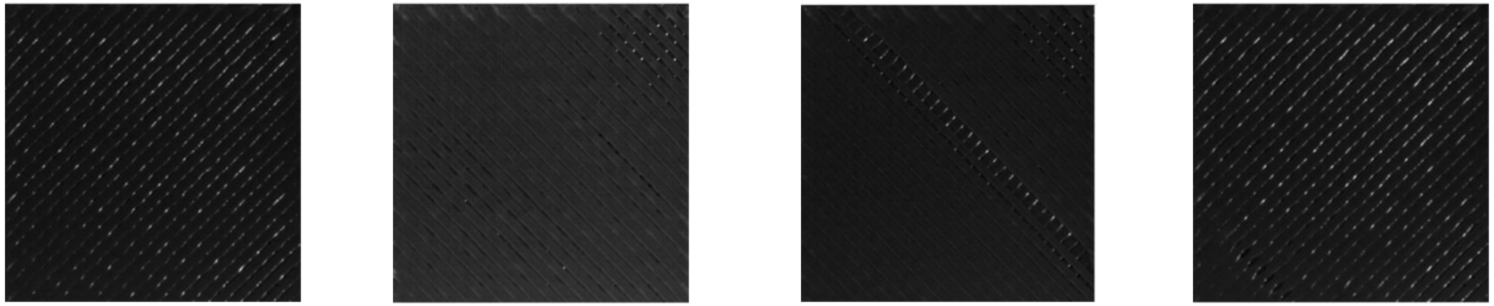


Out of control data 45

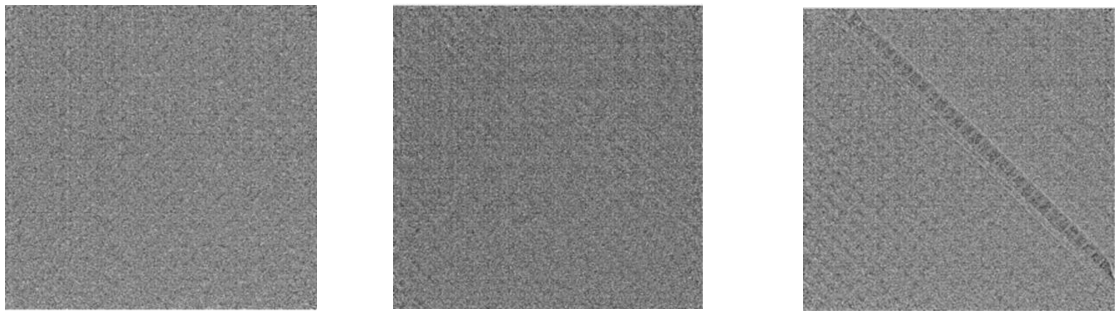
- Current State of the art: they have one model for each pattern
  - (Caltanissetta, Bertoli, & Colosimo) [8]
  - (Bui & Apley) [3]
- We train a single model for both angles

Visual Results (Precise Pixel-Level Segmentation and Image Level Anomaly Detection with Single Model)

Phase2 images



Predicted noise

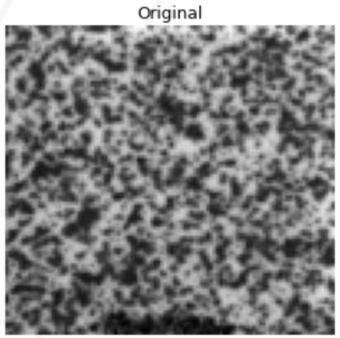


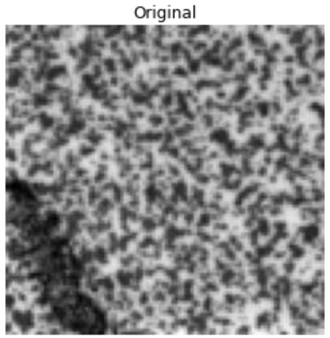
Results LPBF Case Study

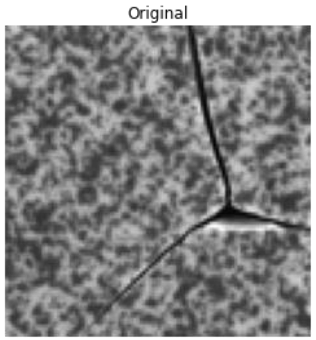
	B&A	C&B	DiffusionAD	AnoDDPM	DDPM	RADAR
Accuracy	0.73	0.67	0.42	0.46	0.5	**0.82**
Precision	0.77	0.70	0.39	0.45	0.6	**0.77**
Recall	0.68	0.64	0.20	0.11	0.14	**0.93**
F1 score	0.72	0.67	0.27	0.18	0.22	**0.85**

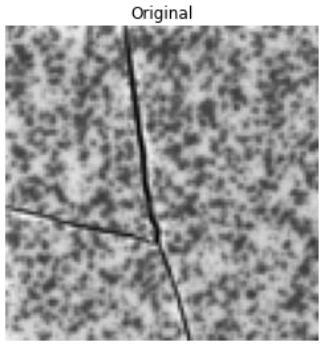
Georgia Tech

Tile Case Study





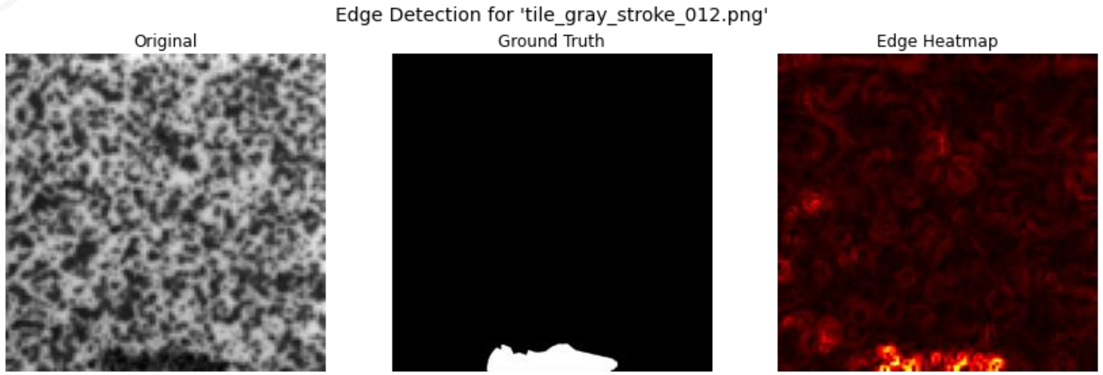


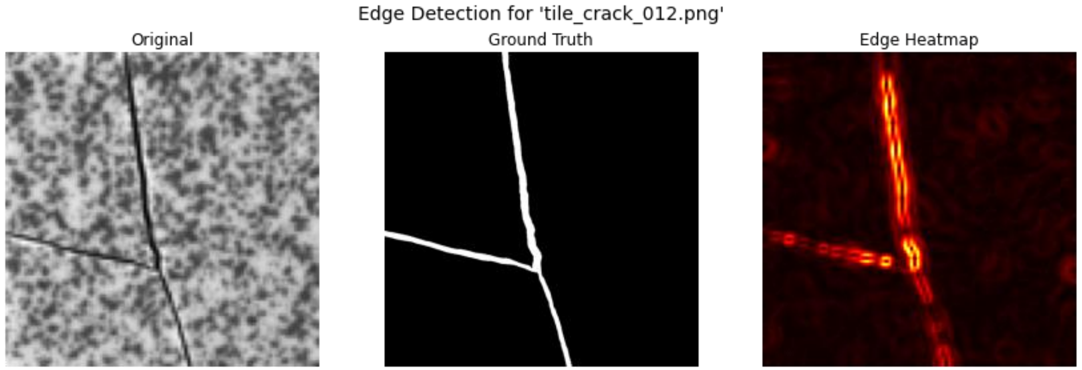


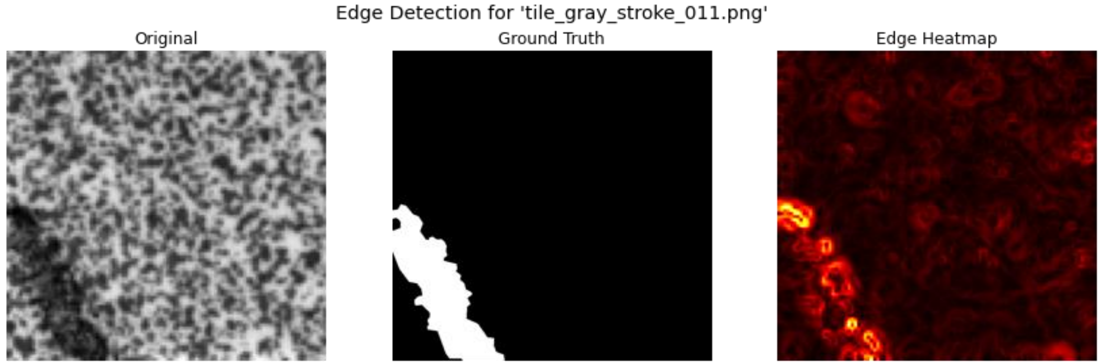
	C&B	B&A	DiffusionAD	AnoDDPM	DDPM	RADAR
Accuracy	0.36	0.51	0.35	0.47	0.43	**0.64**
Precision	1.0	0.87	0.58	0.63	0.59	**0.95**
Recall	0.07	0.35	0.20	0.57	0.57	**0.51**
F1 Score	0.13	0.50	0.30	0.60	0.58	**0.67**

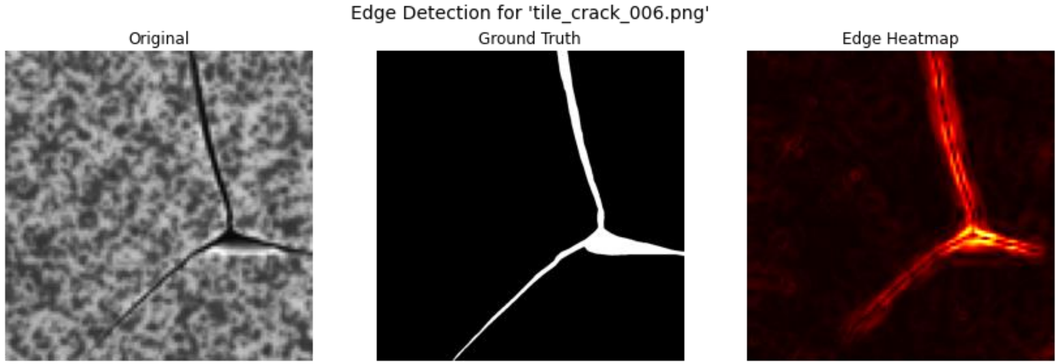
Case Study: Tile

In addition to image level anomaly detection, RADAR shows good pixel level anomaly detection for diagnosis









Ablation Study: Feature Extraction (Contamination=0.05) on the Second Case Study

	Current method	GLCM	SSIM	Hough Transform	Entropy
Accuracy	**0.64**	0.55	**0.59**	0.34	0.39
Precision	**0.95**	0.88	**0.94**	0.80	0.74
Recall	**0.51**	0.41	**0.43**	0.05	0.19
F1 Score	**0.67**	0.56	**0.59**	0.10	0.30

Conclusion

- Generative models show exceptional performance in anomaly detection and segmentation.
  - Current state-of-the-art methods require healthy training data to be effective.
  - Anomalies are localized by reconstructing a normal version of the image through hundreds of sampling steps in the backward diffusion process, followed by residual calculation.
- **Our Contributions:**
  - **RADAR**: Performs diffusion-based anomaly detection and segmentation in a *single step*, unlike reconstruction-based models that require hundreds of steps.
- **Future Work:**
  - Extend the current methods to non-stationary time-series monitoring and anomaly detection:
    - Condition training on past windows to enforce temporal relationship learning.
    - Incorporate a prediction loss to encourage the model to forecast future points.
  - Extend the model to unstructured point-cloud monitoring and anomaly detection:
    - Define patches as neighborhoods of points and train the model to learn their distribution.
    - Develop new feature extraction modules for more precise anomaly localization.

On the Job Market - *Open to Research & ML Opportunities*

Connect on Linkedin

- **Mehrdad Moradi**
  - Focused on robust generative AI for anomaly detection and vision systems
  - Ph.D. Student in Machine Learning, Georgia Tech
  - Advisor: Prof. Kamran Paynabar
  - Research focus: Anomaly Detection, Diffusion Models
- **Selected Publications**
  - **Moradi, M.**, Chen, S., Yan, H., Paynabar, K. A Single Image Is All You Need: Zero-Shot Anomaly Localization Without Training Data. (Submitted to WACV 2026) [9]
  - **Moradi, M.**, Grasso, M., Colosimo, B. M., Paynabar, K. Single-Step Reconstruction-Free Anomaly Detection and Segmentation via Diffusion Models. (ICMLA 2025) [7]
  - **Moradi, M.**, Paynabar, K. RDDPM: Robust Denoising Diffusion Probabilistic Model for Unsupervised Anomaly Segmentation. (ICCVW 2025) [8]
- **Opportunities**
  - Actively seeking Machine Learning research or applied roles (internship or full-time) starting in Spring, Summer, or Fall 2026.
- **Contact:**
  - Email: mmoradi6@gatech.edu

References

- [1] Candès, Emmanuel J., et al. "Robust principal component analysis?." *Journal of the ACM (JACM)* 58.3 (2011): 1–37.
- [2] Yan, Hao, Kamran Paynabar, and Jianjun Shi. "Anomaly detection in images with smooth background via smooth-sparse decomposition." *Technometrics* 59.1 (2017): 102–114.
- [3] Bui, Anh Tuan, and Daniel W. Apley. "A monitoring and diagnostic approach for stochastic textured surfaces." *Technometrics* 60.1 (2018): 1–13.
- [4] Buerhop-Lutz, Claudia, et al. "A benchmark for visual identification of defective solar cells in electroluminescence imagery." *35th European PV Solar Energy Conference and Exhibition*. Vol. 12871289. 2018.
- [5] Bergmann, Paul, et al. "MVTec AD – A comprehensive real-world dataset for unsupervised anomaly detection." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019.
- [6] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. "Denoising diffusion probabilistic models." *Advances in Neural Information Processing Systems* 33 (2020): 6840–6851.
- [7] Wyatt, Julian, et al. "Anoddpm: Anomaly detection with denoising diffusion probabilistic models using simplex noise." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- [8] Caltanissetta, Fabio, Luisa Bertoli, and Bianca Maria Colosimo. "In-situ monitoring of image texturing via random forests and clustering with applications to additive manufacturing." *IISE Transactions* 56.10 (2024): 1070–1084.

References

- [9] Zhang, Hui, et al. “DiffusionAD: Norm-guided one-step denoising diffusion for anomaly detection.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).
- [10]: Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, and Gian Luca Foresti. Vt-adl: A vision transformer network for image anomaly detection and localization. In 2021 IEEE 30th International Symposium on Industrial Electronics (ISIE), pages 01–06. IEEE, 2021. 5, 7, 8, 4
- [11]: M. -J. Wu, J. -S. R. Jang and J. -L. Chen, "Wafer Map Failure Pattern Recognition and Similarity Ranking for Large-Scale Data Sets," in IEEE Transactions on Semiconductor Manufacturing, vol. 28, no. 1, pp. 1-12, Feb. 2015, doi: 10.1109/TSM.2014.2364237
- [12]: Moradi, Mehrdad, et al. "Single-Step Reconstruction-Free Anomaly Detection and Segmentation via Diffusion Models." *arXiv preprint arXiv:2508.04818* (2025).
- [13]: Moradi, Mehrdad, and Kamran Paynabar. "RDDPM: Robust Denoising Diffusion Probabilistic Model for Unsupervised Anomaly Segmentation." *arXiv preprint arXiv:2508.02903* (2025).

Georgia Tech