RDDPM: Robust Denoising Diffusion Probabilistic Model for Unsupervised Anomaly Segmentation

Improving diffusion model robustness under contaminated training data

Mehrdad Moradi[1]

Kamran Paynabar[1]

1. H. Milton Stewart School of Industrial and Systems Eng. And ML center, Georgia Institute of Technology

3rd Workshop on Vision-based Industrial Inspection, ICCV 2025,
Honolulu, Hawaii, October 2025

Georgia Tech

Outline

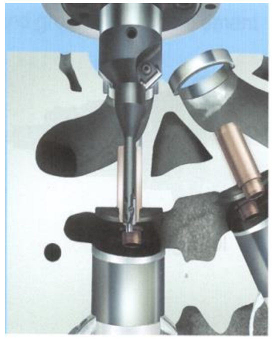
- Motivation
- Research Gap
- Problem Formulation
- Proposed **Robust Diffusion Training Algorithm**
- Quantitative and Qualitative Results
- Sensitivity Analysis

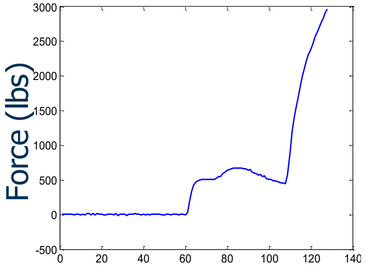
Paper Link

Georgia Tech

Motivation

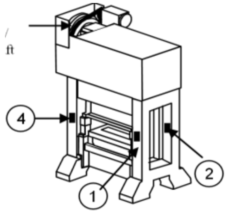
Valve Seat Assembly

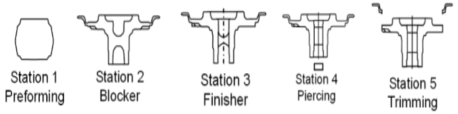


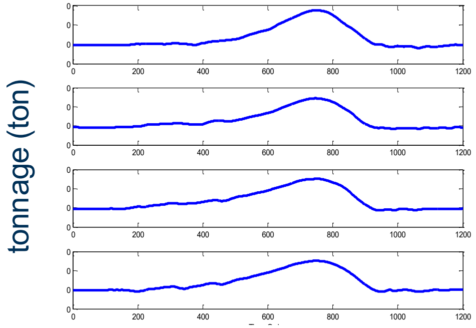


Time of each operation cycle

Forging

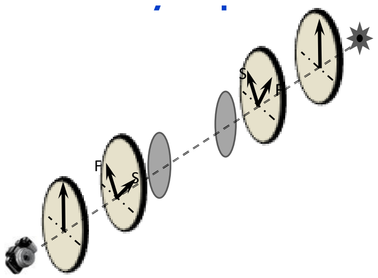


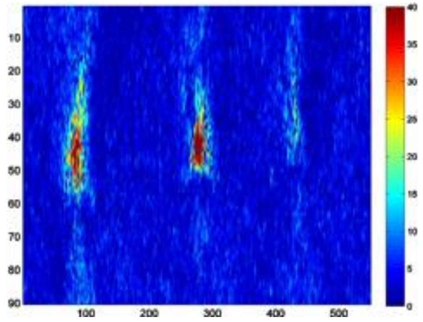




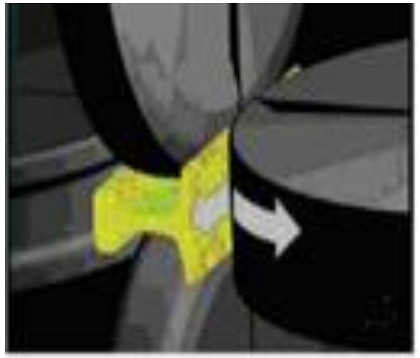
Crank angle (degree)

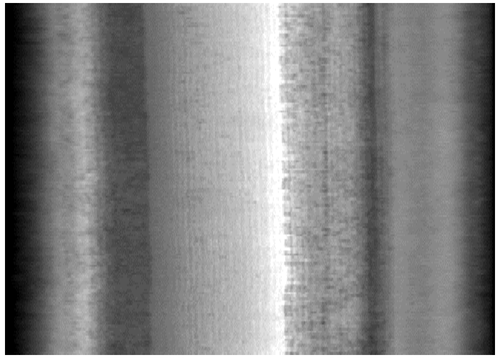
Semiconductor





Rolling





Anomaly Detection in HD Data has wide applications in different domains. We aim to detect subtle defects across heterogeneous high-dimensional signals and textures.

Research gap

Classical statistical methods rely on restrictive assumptions

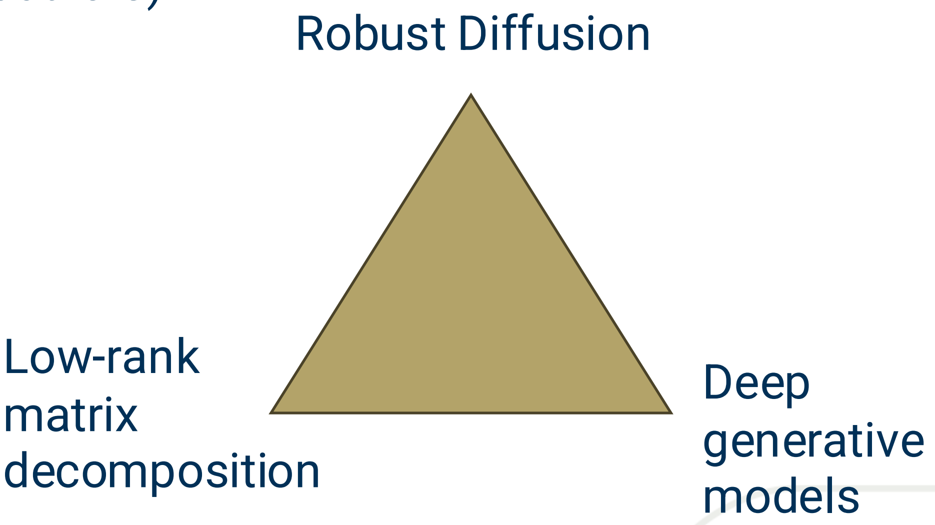
- RPCA [1]: relies on a low-rank assumption on the background and sparsity of the anomaly
- SSD [2]: relies on the smoothness of the normal background and the sparsity of the anomaly
- However, they fail on complex, high-dimensional data.

Deep generative approaches rely on supervised assumptions:

- Reconstruction-based (VAE [10], GAN [11], Diffusion [6]) anomaly detection needs healthy data for training

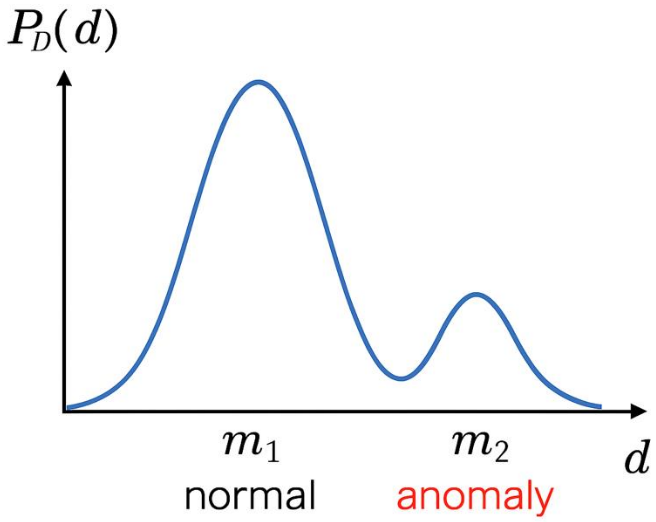
We propose Robust **DDPMs:**

- Do not assume low rank (linear LD space)
- Do not need healthy training data
- The only assumption is that the probability of having an anomalous sample is the training data is significantly lower than in healthy data (i.e., outliers)



Problem Formulation

- Let $\{d_1, d_2, \ldots, d_n\}$ be the set of training observations coming from an unknown distribution $P_D(d)$
- P has two modes: $m1, m2$ corresponding to normal and anomaly, respectively.
- Our objective is to decompose the new anomalous image into normal and anomalous component
- $Y = n + a \quad \mathrm{s.t.}\ \mathrm{n} \sim P_{m_1}$: normal mode of the distributio$n$
- $n \sim p(x_0|x_{t_0}), t_0 < T, x_{t_0} \sim q(x_{t_0}|y)$
- $q(x_{t_0}|y)$: predefined forward conditional distribution
- $p(x_0|x_{t_0})$: learned backward conditional distribution
- T: *number of diffusion timesteps*

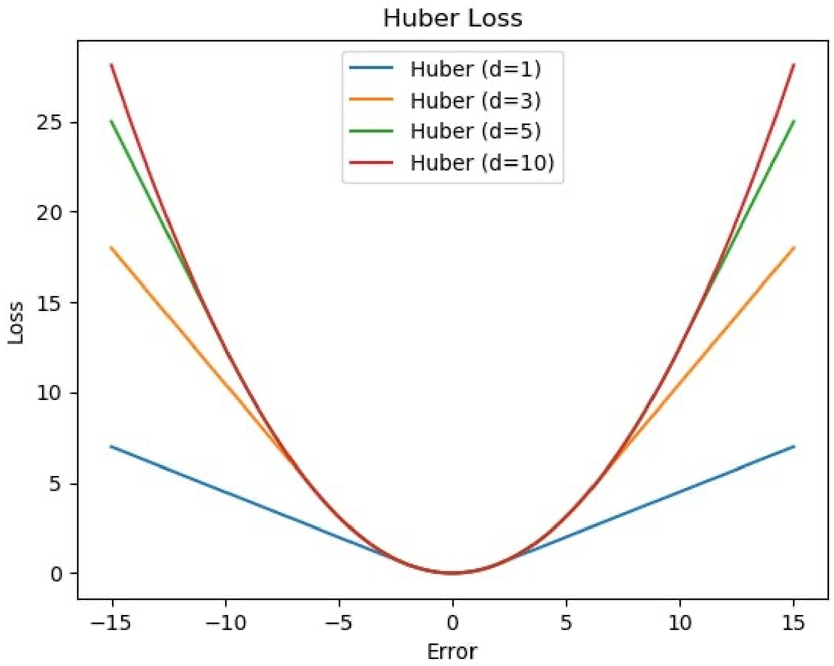


Two-mode data distribution assumption

Huber vs MSE loss

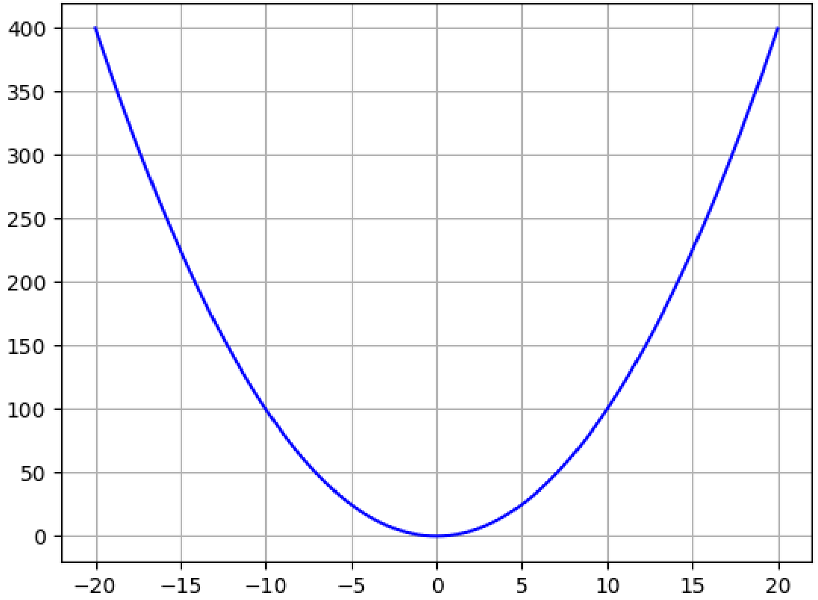
$$L_\delta(a) = \begin{cases} \frac{1}{2}a^2 & \text{for } |a| \le \delta, \\ \delta \cdot \left(|a| - \frac{1}{2}\delta\right), & \text{otherwise.} \end{cases}$$

a, δ: model residual, hyperparameter controlling robustness



$$L(a) = a^2$$

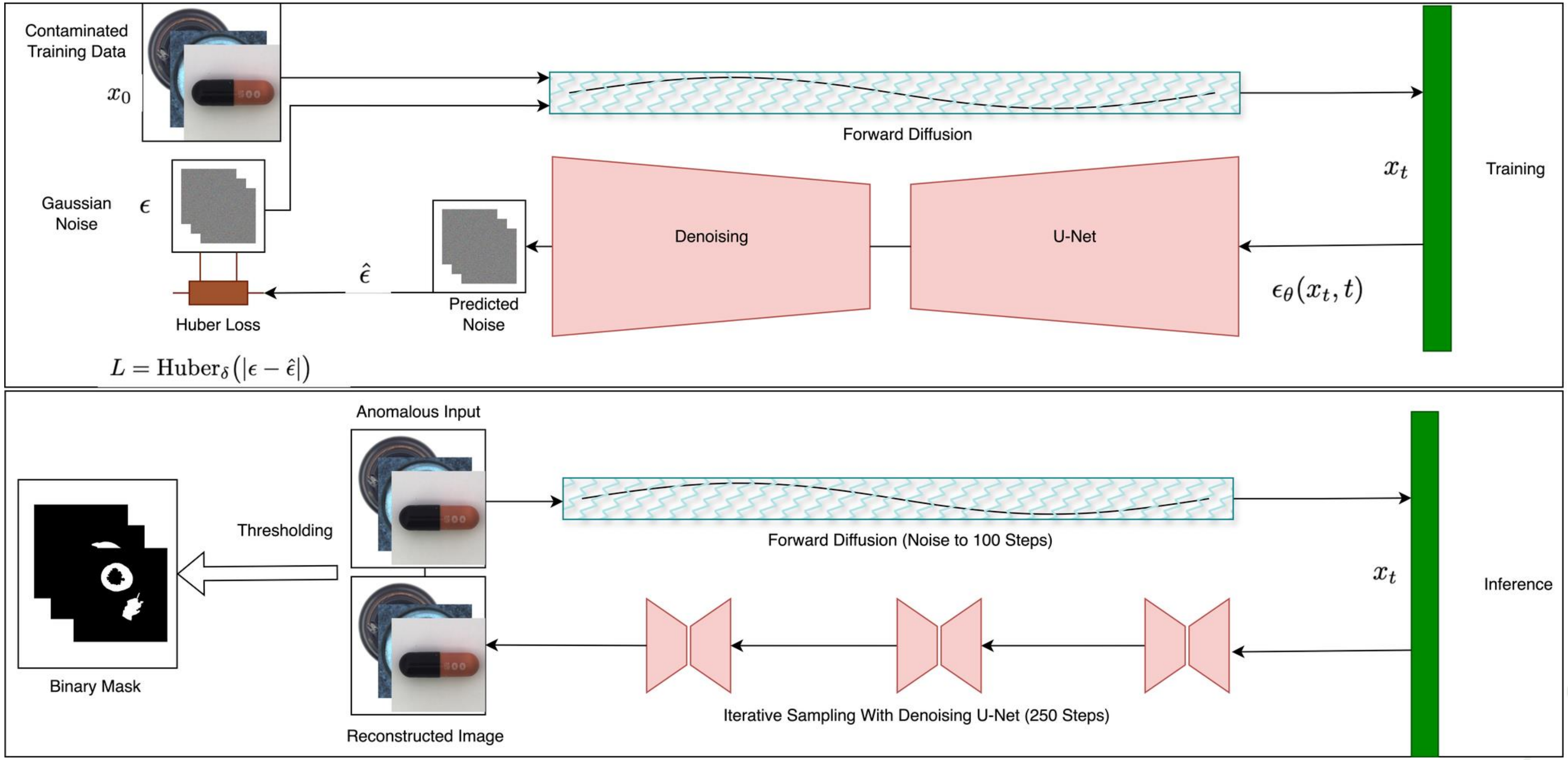
a, δ: model residual



Huber offers resilience to outliers — a key advantage for contaminated data

Proposed Robust Diffusion Model

- We use Huber loss to make DDPM robust against outliers and propose a new training algorithm.
- We add noise through 100 forward diffusion steps and denoise for 250 steps to reconstruct the healthy image.
- This pipeline allows training directly on contaminated datasets.



Robust Diffusion Training Algorithms

- RDDPM-Huber: trained with Huber loss penalizing larger residuals with L1 norm
- RDDPM-LTS: using least trimmed squares loss, keeping the top s samples with the lowest residuals
- We use RDDPM-Huber in our experiments because it performed better empirically

RDDPM-LTS

while Not converged **do**

$x_0 \sim q(x_0)$

$t \sim \text{Uniform}(\{1, \ldots, T\})$

$\epsilon \sim \mathcal{N}(0, I)$

Take gradient descent step on

$$\nabla_\theta LTS(\left\| \epsilon - \epsilon_\theta \left(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right\|^2)$$

$$= \sum_{i=1}^{s=\lambda \times B} \nabla_\theta \left\| \epsilon_i - \epsilon_\theta \left(\sqrt{\bar{\alpha}_{t_i}} x_{0_i} + \sqrt{1 - \bar{\alpha}_{t_i}} \epsilon_i, t_i \right) \right\|^2$$

Where $s \in \{1, ..., B\}$ and $\lambda \in (0, 1]$

end while

RDDPM-Huber

while Not converged **do**

$x_0 \sim q(x_0)$

$t \sim \text{Uniform}(\{1, \ldots, T\})$

$\epsilon \sim \mathcal{N}(0, I)$

Take gradient descent step on

$$\nabla_\theta \, \text{Huber}_\delta \left(\epsilon - \epsilon_\theta \left(\sqrt{\bar{\alpha}_t} x_0 + \sqrt{1 - \bar{\alpha}_t} \epsilon, t \right) \right)$$

$$\text{where} \quad \text{Huber}_\delta(r) = \begin{cases} \frac{1}{2} r^2 & \text{if } |r| \leq \delta \\ \delta \left(|r| - \frac{1}{2} \delta \right) & \text{if } |r| > \delta \end{cases}$$

end while

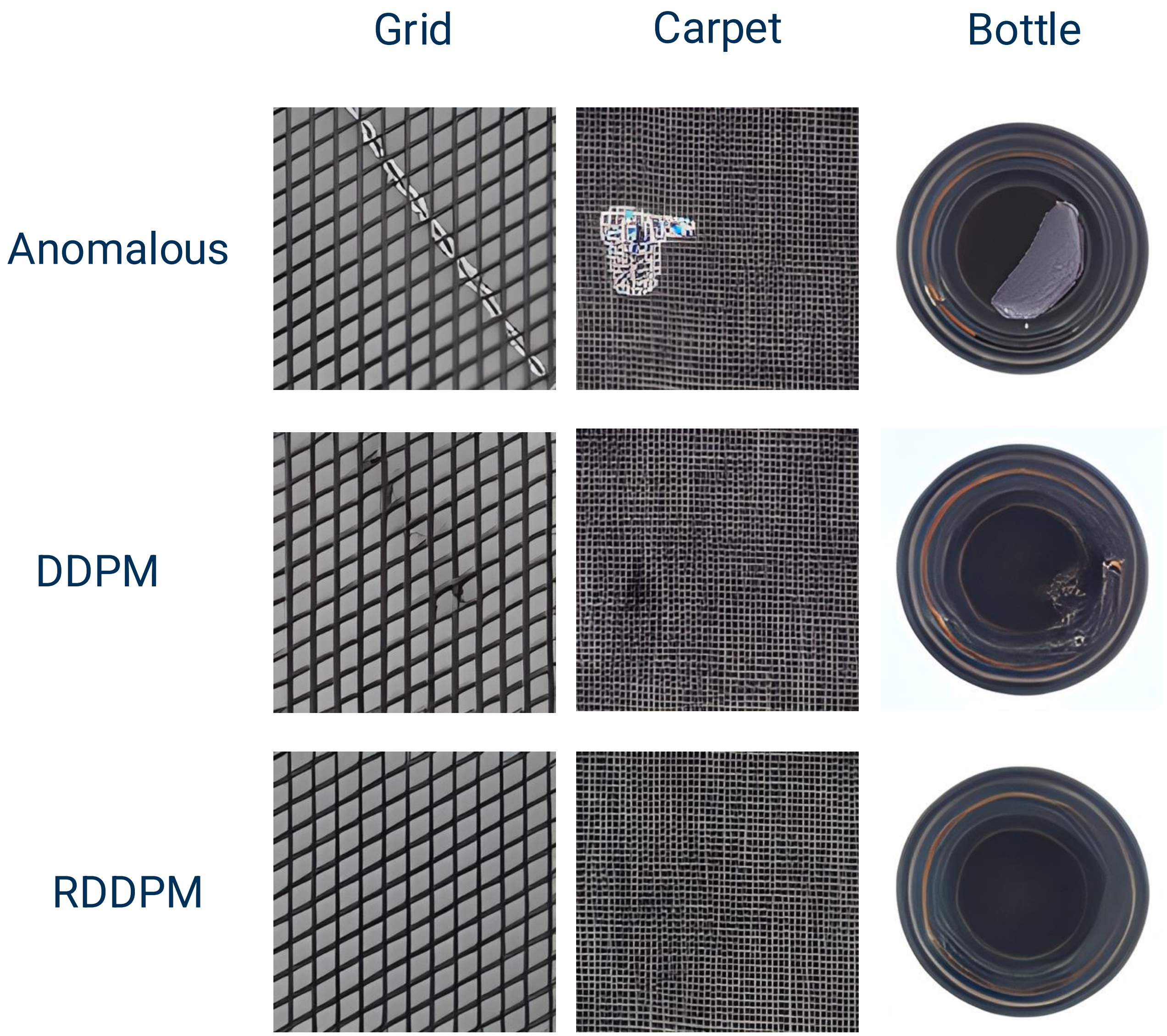
Results

- Trained on 20% contamination, reconstructions by RDDPM are cleaner than DDPM.
- RDDPM outperforms other diffusion models on Carpet, Grid, and the entire MVTec-AD [3] dataset.

Carpet	AUROC	AUPRC	MSE
RDDPM	**0.5673**	**0.0362**	0.1246
AnoDDPM [5]	0.4650	0.0234	0.2115
DiffusionAD [6]	0.4909	0.0268	**0.1199**

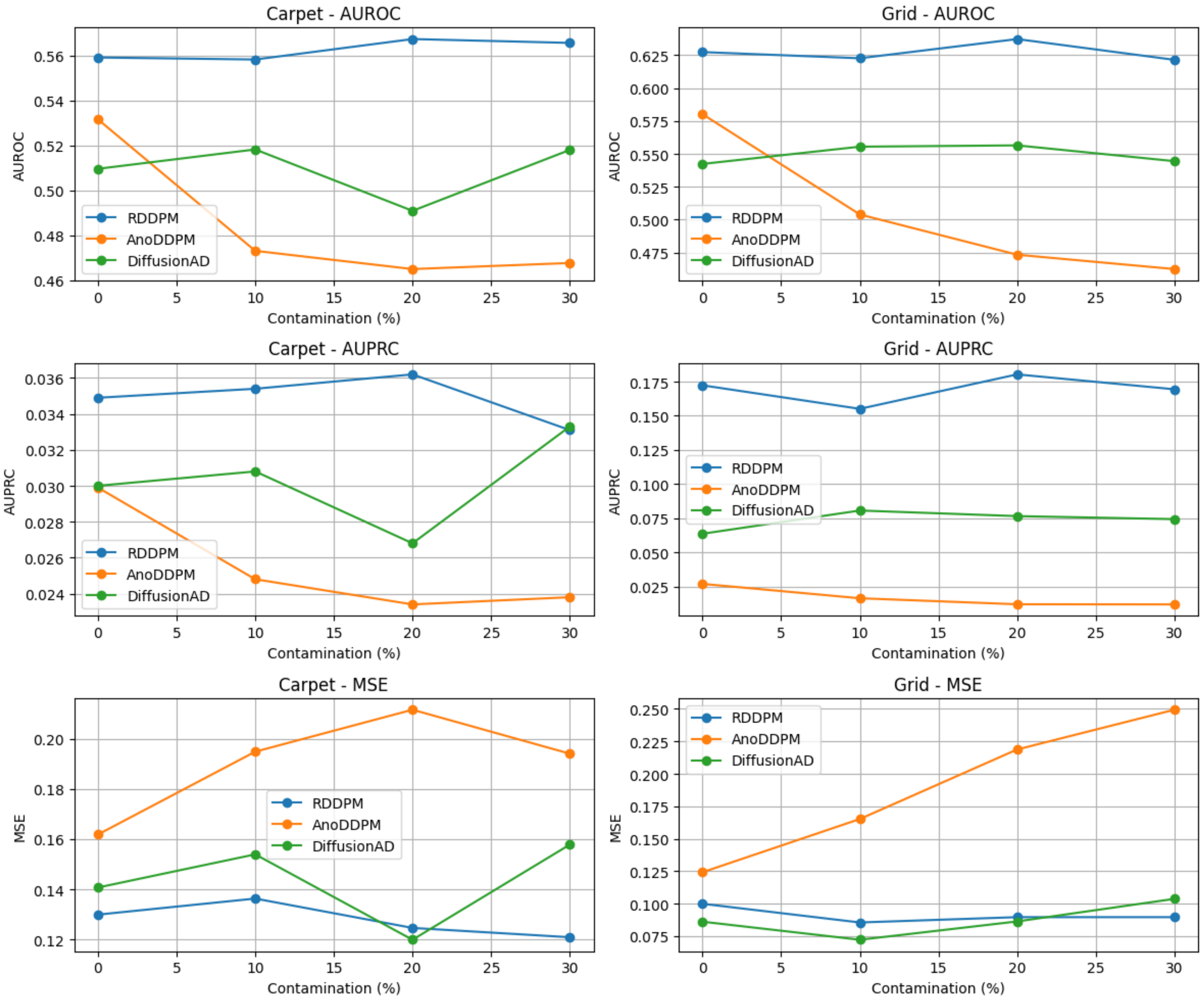
Grid	AUROC	AUPRC	MSE
RDDPM	**0.6373**	**0.1803**	0.0896
AnoDDPM	0.4734	0.0121	0.2188
DiffusionAD	0.5565	0.0766	**0.0863**

MVTec-AD	AUROC-ID	AUROC-OOD
RDDPM	**0.78**	**0.71**
DDPM	0.76	0.69



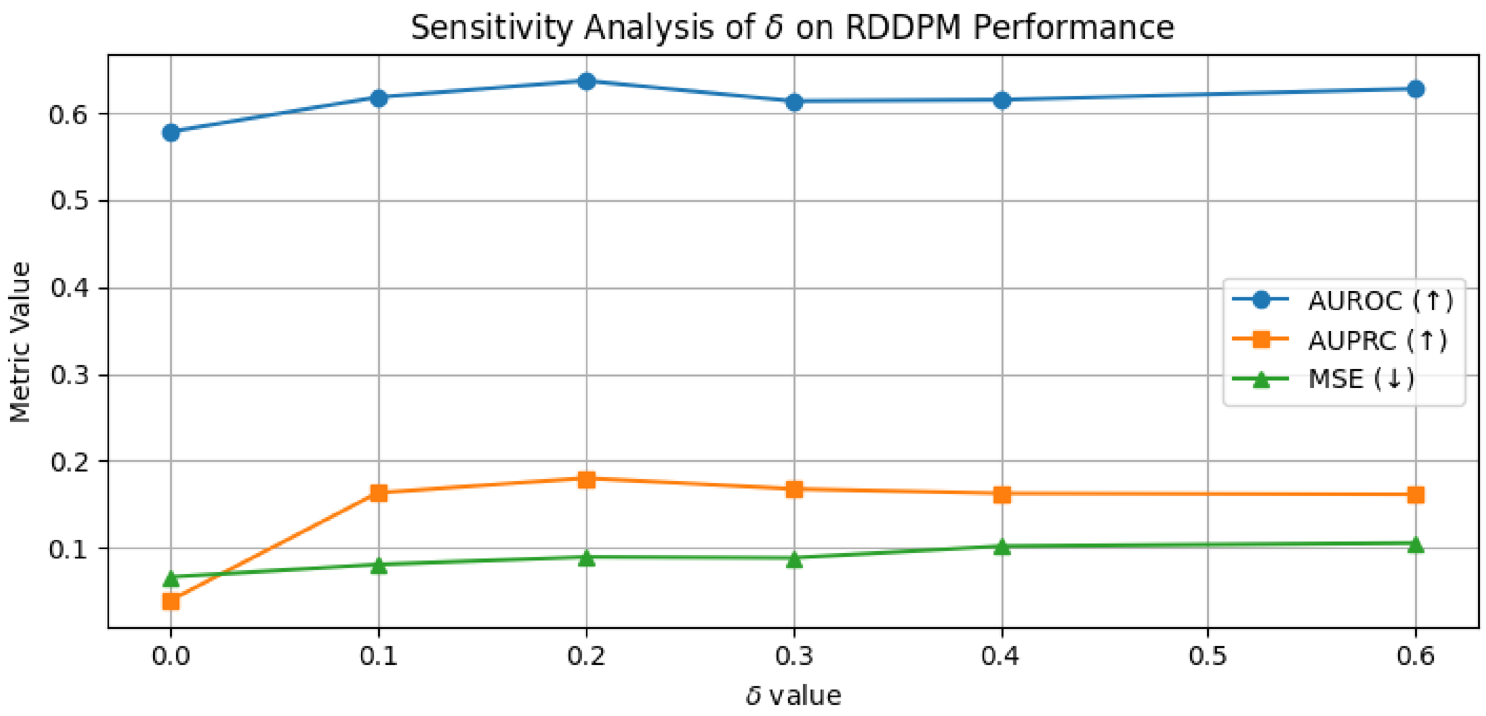
Sensitivity Analysis: Contamination Level

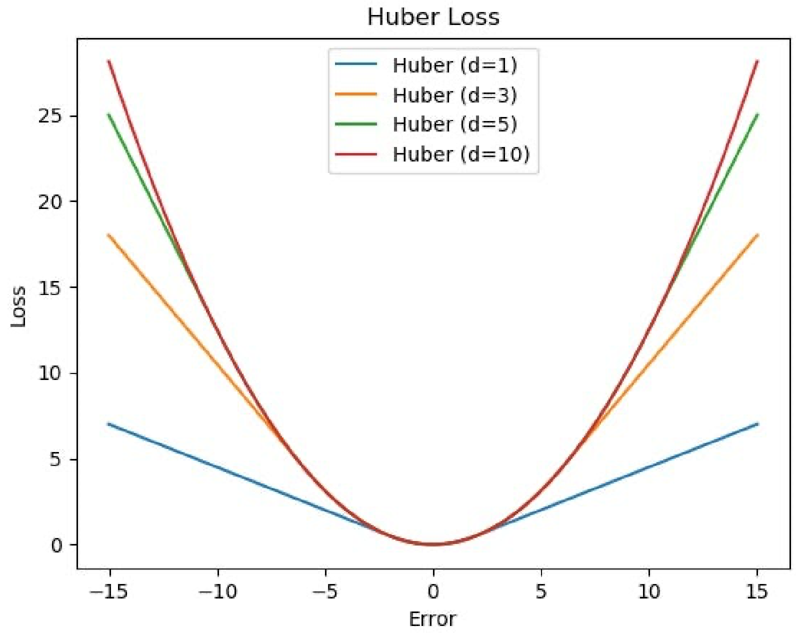
- RDDPM outperforms other methods across different contamination levels.
- In zero contamination, it shows a better performance in AUROC and AUPRC.
- RDDPM maintains stable performance even under high contamination.



Robustness Parameter

- *When δ = 0 → equivalent to L1 norm (poor performance)*
- *When δ → ∞ → equivalent to L2 norm (DDPM formulation)*
- *Overall, performance is largely insensitive to δ variations.*





Conclusion & Future Directions

- Generative diffusion models are highly effective for anomaly detection and segmentation.
- However, most existing approaches rely on clean training data, which is unrealistic in real-world industrial settings.
- Our proposed **RDDPM** relaxes this assumption, performing robustly on contaminated data while maintaining strong detection accuracy.
- Future research:
  - Extending **RDDPM** to *unstructured point-cloud data* for 3D anomaly detection.
  - Adapting **RDDPM** to *non-stationary time-series signals* for temporal anomaly detection.
  - Generalizing the framework to an *extensive family of robust loss functions*, forming a family of **Robust Diffusion Models**.

Thank you! Questions are welcome.

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] 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.
- [4] Ho, Jonathan, Ajay Jain, and Pieter Abbeel. “Denoising diffusion probabilistic models.” *Advances in Neural Information Processing Systems* 33 (2020): 6840–6851.
- [5] 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.
- [6] Zhang, Hui, et al. “DiffusionAD: Norm-guided one-step denoising diffusion for anomaly detection.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2025).
- [7]: Moradi, Mehrdad, et al. "Single-Step Reconstruction-Free Anomaly Detection and Segmentation via Diffusion Models." *arXiv preprint arXiv:2508.04818* (2025).
- [8]: Moradi, Mehrdad, and Kamran Paynabar. "RDDPM: Robust Denoising Diffusion Probabilistic Model for Unsupervised Anomaly Segmentation." *arXiv preprint arXiv:2508.02903* (2025).
- [9]: Moradi, Mehrdad, Shilin Chen, Huan Yan, and Kamran Paynabar. “A Single Image Is All You Need: Zero-Shot Anomaly Localization Without Training Data.” *arXiv preprint arXiv:2508.07316* (2025).

References

- [10]: Kingma, Diederik P., and Max Welling. “Auto-encoding variational Bayes.” *arXiv preprint arXiv:1312.6114* (2013).
- [11]: Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative adversarial nets.” *Advances in Neural Information Processing Systems* 27 (2014).