Conditional Diffusion Models for OOD Detection

Author: Ahmed Mohammed

Institute for Machine Learning – Supervisor: Sepp Hochreiter, Claus Hofmann

JOHANNES KEPLER UNIVERSITY LINZ

Introduction & Motivation

Problem: Traditional classifiers struggle to recognize out-of-distribution (OOD) samples, leading to inaccurate and overconfident predictions on unseen data

Motivation: Conditional diffusion models can learn rich data representations and measure reconstruction quality

Approach: Leverage reconstruction error principle - in-distribution samples reconstruct better than OOD samples

Key Insight: Class-conditional generation enables precise anomaly detection through comparative reconstruction errors

Methodology Details

Diffusion Model Architecture:

- UNet2D is conditioned on class labels
- Sample size: 32×32, Embedding dim: 128
- Block channels: [128, 256, 512, 512]
- DDPM scheduler with linear β-schedule

Training Loss: Reconstruction loss between predicted and actual noise

$$\boldsymbol{L} = \left|\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}_{\boldsymbol{\theta}(\boldsymbol{x_t},\boldsymbol{t},\boldsymbol{c})}\right|^2$$

OOD Detection Process [1]**:**

1. For test image x, run multiple trials (10 iterations)
2. Add noise: $x_t = \sqrt{\bar{\alpha}_t}\, x + \sqrt{(1-\bar{\alpha})}\, \epsilon$
3. Predict noise for each class: $\epsilon_\theta = f\_\theta(x_t, t, c)$
4. Compute MSE error per class
5. Classification via softmax over negative errors

Notations:

Symbol	Meaning
x	Original clean image
x_t	Noisy image at time step t
ϵ	Gaussian noise, $N(0, I)$
ϵ_θ	Predicted noise at each timestep
$\bar{\alpha}_t$	Cumulative noise coefficient

Experiment Setup

Dataset: CIFAR-10

- **In-Distribution**: Airplane class (Class 0)
- **Out-of-Distribution**: All other 9 classes (Class 1)
- **Training**: 5,000 airplane images + 5,000 mixed other classes
- **Validation**: 1,000 airplane + 1,000 other classes

Training Configuration:

- Batch size: 32 | Epochs: 500
- Optimizer: AdamW (lr=1e-4, wd=1e-4)
- Scheduler: Cosine annealing
- Sampling: Weighted (class balanced)

References

[1] Bowen Li, Robin Rombach, Vladlen Koltun, and Luke Zettlemoyer. *Your Diffusion Model is Secretly a Zero-Shot Classifier.* In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
[2] Andreas Kirsch, Joost van Amersfoort, and Yarin Gal. *Denoising Diffusion Models for Out-of-Distribution Detection.* In Proceedings of the 40th International Conference on Machine Learning (ICML), 2023.
[3] Daniel Smolyar, Nicholas Carlini, and Dhruv Madeka. *Intriguing Properties of Generative Classifiers.* In Proceedings of the 41st International Conference on Machine Learning (ICML), 2024.

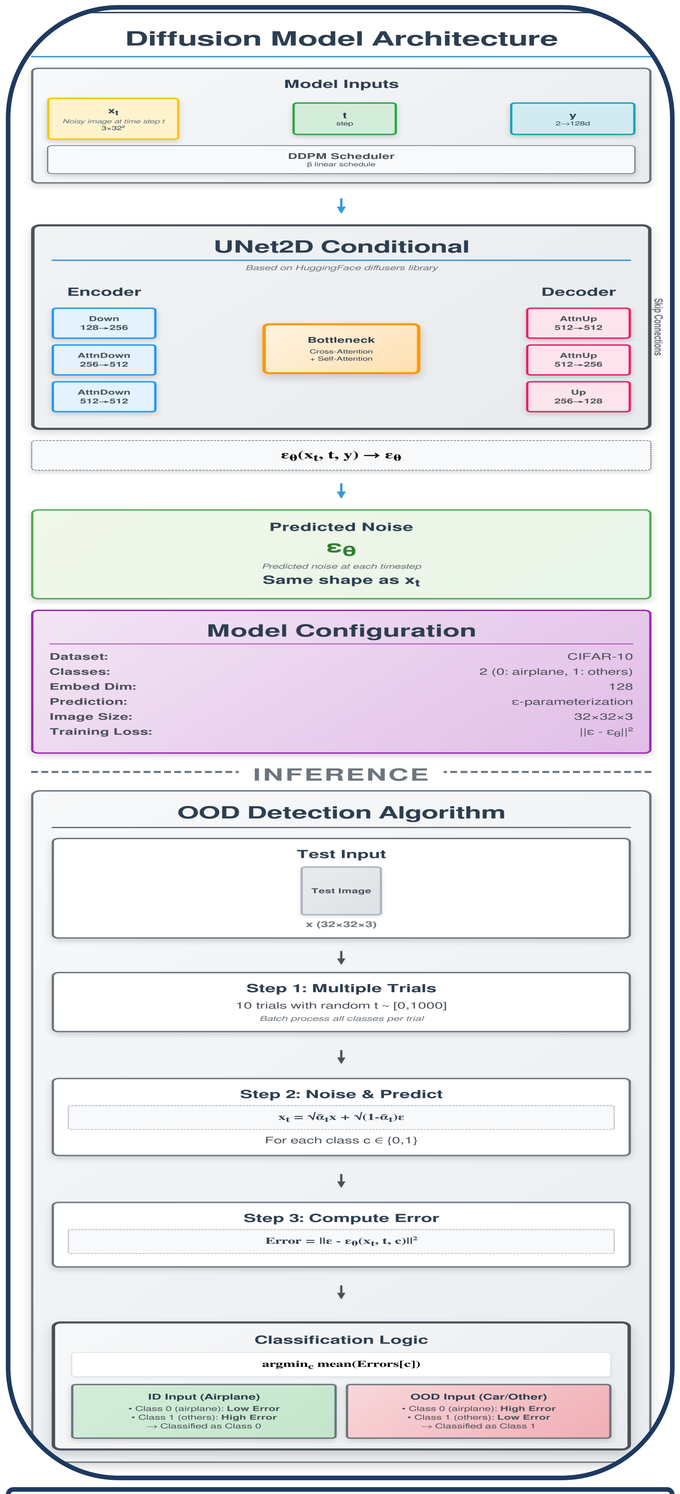

Figure 1: OOD Detection Pipeline Architecture

Results

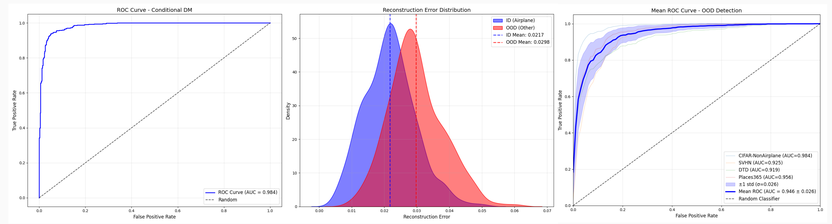

Figure 2: (Left) ROC curve and (Middle) reconstruction error distribution for the conditional diffusion model trained on CIFAR-10 (ID: Airplane, OOD: Other). (Right) Mean ROC curve across multiple OOD datasets with shaded area representing ±1 std. deviation.

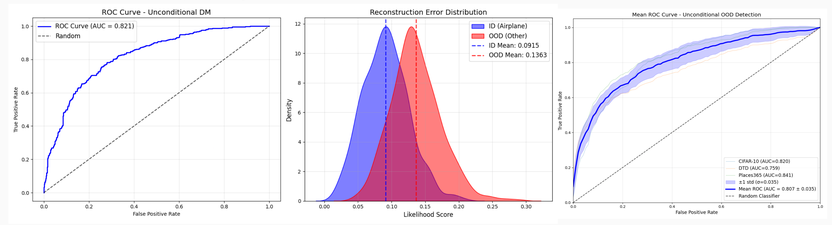

Figure 3: (Left) ROC curve and (Middle) reconstruction error distribution for the unconditional diffusion model trained on CIFAR-10 (ID: Airplane). (Right) Mean ROC curve across multiple OOD datasets with shaded area representing ±1 std. deviation.

Performance Comparison

OOD Datasets	Conditional Diffusion (AUROC%)	Unconditional Diffusion (AUROC%)
Cifar-10 (Airplane VS Non-Airplane)	98.40	82.10
Places365	95.60	84.10
SVHN	92.50	61.60
DTD	91.90	75.90

Conclusion & Future Work

Conclusion:

- Conditional diffusion models achieve significantly higher AUROC scores for OOD detection (98.40 vs 82.10)
- Reconstruction error provides reliable anomaly scoring
- Method generalizes across multiple datasets (CIFAR-10, SVHN, etc.)
- Class conditioning significantly improves detection accuracy

Future Work:

- Extend to multi-class OOD detection beyond binary classification
- Investigate other noise schedules and sampling strategies
- Apply to high-resolution datasets and real-world applications
- Develop computational efficiency improvements for deployment
- Explore uncertainty quantification in reconstruction errors