

DIFFCONT: CONTINUAL ANOMALY DETECTION WITH DIFFUSION MODELS AND OUTLIER REJECTION

Anonymous authors

Paper under double-blind review

ABSTRACT

We address continual anomaly detection in industrial inspection where new data flows in periodically in the form on new categories or newer types of defective samples. In such a setting, we would want the model to generalize to newer categories and shifting data distributions whilst maintaining its performance on previously learnt categories. Building on a diffusion-based U-Net with conditional generation and an auxiliary classifier, we introduce a Weibull-guided pipeline that serves two roles: (i) as a representative conditional generative distillation model, where a per-class Weibull model is fit on a projected embedding space and accepts only strong exemplar replay samples via threshold; (ii) as a per-class anomaly detection model, where the same heavy-tailed Weibull fit on the tail of normal embeddings converts distances into outlier probabilities for open-set anomaly detection. On MVTec-AD dataset split into 15 class-incremental experiences, our method achieves superior performance across experiences while presenting a light-weight anomaly detection workflow fit for industrial use.

1 INTRODUCTION

A visual system is subjected to constantly evolving data streams with a myriad of irregularities: categories appearing, tooling changes, sensors drift - and yet we expect yesterday’s notion of “normal” to hold tomorrow. This is made incredibly clear in industrial inspection, represented well by MVTec-AD’s 15 categories with normal-only training and diverse, previously unseen defects at test time. Here, adaptability is key to overcome catastrophic forgetting as new categories arrive Kirkpatrick et al. (2017). Naive fine-tuning is prone to such forgetting. Continual Learning (CL) is therefore necessary to trade plasticity for stability under realistic compute and memory budgets. Traditional CL approaches regularise (e.g. Elastic Weight Consolidation (EWC) Kirkpatrick et al. (2017)), replay exemplars (e.g., iCarl Rebuffi et al. (2017)) or synthesise data (e.g. generative replay Shin et al. (2017)), yet all of these fail under anomaly detection conditions as opposed to the more oft considered classification.

Generative replay is attractive in the continual anomaly detection setting, especially with recent diffusion models offering high fidelity, controllable generators Ho et al. (2020); Song et al. (2020b), as it avoids storing raw data. However, replay alone is not enough: generated samples can drift away from their true class manifolds, degrading both representation learning and the calibration of threshold scores for anomalies. At the same time, anomaly detection can be treated as an open-set problem Mundt et al. (2022). Combining these together, we obtain generative distillation with outlier rejection of out-of-distribution samples, thereby preserving the entire reverse process across experiences, making diffusion replay viable at scale. Even with this strengthened replay, anomaly detection can essentially be treated as Open-set recognition and its Extreme Value Theory (EVT) dictate that we should explicitly model the tails, so that the system can reject the unknown rather than force a closed-set guess.

We embrace an experience-centric view by coupling diffusion-based CL with explicit tail modelling on a per-class basis. More concretely, we train a diffusion U-Net guided by an auxiliary classifier and a 1x1-convolutional projector at the bottleneck to define a stable, low-dimensional embedding where experience-to-experience comparisons are meaningful. Within this space, we fit Weibull models for every class to the right tail of the normalised embeddings. These models play a dual role in the training loop. First, they screen samples generated for generative distillation - rejecting outliers

and ensuring that clean, representative samples are replayed. Secondly, at test time, the distances from the class centre are converted into outlier probabilities based on the CDF of the heavy-tailed Weibull distribution for that class which are then thresholded to classify the samples as 'normal' or 'anomaly'.

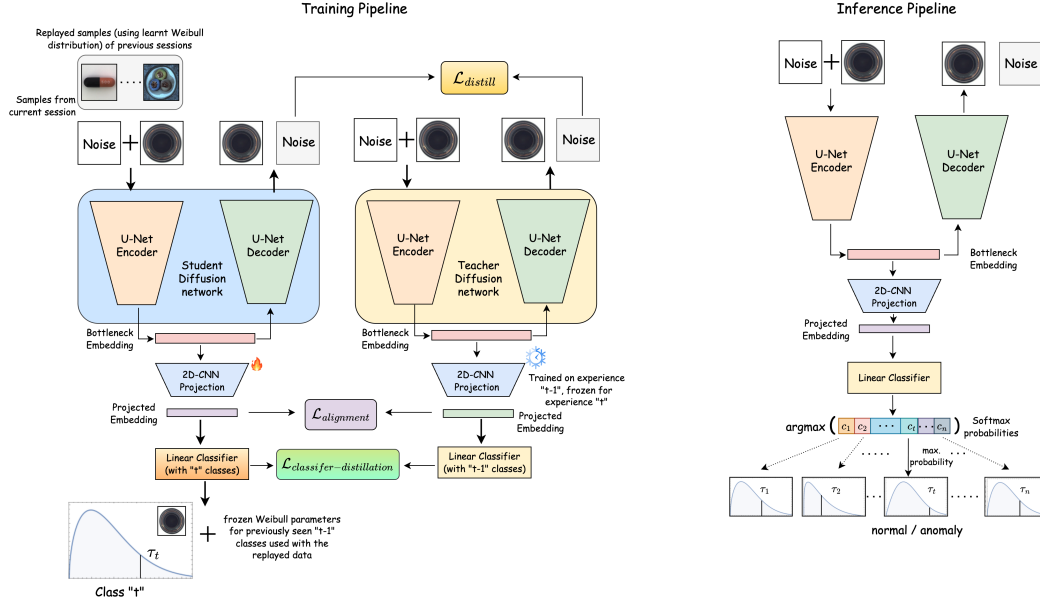


Figure 1: DiffCont Continual Anomaly Detection Workflow. First a noised image is passed to the diffusion U-Net. Secondly, a projected embedding space is trained on the first experience using a 1×1 Conv2D layer (this is later frozen). Then a linear classifier is trained on the output of the projection space, P . Next, we perform generative distillation between experiences using a typical teacher-student setup. In this scenario, the generated images must be below a per-class Weibull threshold such that we are replaying representative samples between experiences. Finally, the samples are classified as normal or anomalous depending on the output of the CDF of a class-conditioned Weibull distribution being below a particular threshold, which is determined using a validation set. During inference, projected embeddings are extracted from the U-Net Bottleneck + Projector P and classified by the Classifier Head H into class c . We select the appropriate Weibull distribution (as was fit for class c during training) to obtain the cumulative probability for the sample which is thresholded by τ_c into 'normal' or 'anomaly' class.

2 RELATED WORK

2.1 DIFFUSION MODELS

Diffusion models have become a cornerstone in generative modelling, especially for high-quality image synthesis. These models learn to gradually transforming noise into data samples through a denoising process, effectively modelling complex distributions. Denoising Diffusion Implicit Models (DDIM) improves sample quality and speed using non-Markovian diffusion processes Song et al. (2020a). Latent Diffusion Models (LDM) further accelerate generation by operating in a lower-dimensional latent space while preserving high fidelity Rombach et al. (2022). Model distillation techniques have been applied to diffusion models to reduce sampling steps, enabling faster and more efficient inference without significant quality loss Salimans & Ho (2022).

2.2 CONTINUAL LEARNING WITH REPLAY METHODS

Continual learning aims to mitigate catastrophic forgetting when models learn from a stream of tasks. Replay-based methods store and rehearse past examples to maintain performance. Experience Replay (ER) is a foundational approach that intermittently retrains on buffered samples from

previous tasks Rolnick et al. (2019) DER++ extends replay by combining rehearsal with knowledge distillation to align feature representations across tasks, yielding improved stability and plasticity Buzzega et al. (2020).

2.3 OUTLIER REJECTION

Ensuring robust model performance under distributional shifts in Continual Learning requires effective mechanisms for outlier rejection. Mundt et al. (2019) proposed a unified probabilistic framework for outlier rejection leveraging likelihoods under learned model distributions based on variational inference in a single deep autoencoder model. This formulation enables principled detection of novel or corrupted inputs by evaluating the joint likelihood of observations, surpassing heuristic or threshold-based anomaly rejection strategies. The approach constrains the approximate posterior by identifying high-density regions associated with correctly classified samples. These bounds serve a dual role: they facilitate reliable separation of unseen, out-of-distribution inputs from previously learned tasks, and they guide generative replay by restricting synthesis to strictly in-distribution data, thereby mitigating catastrophic interference.

2.4 ANOMALY DETECTION ON MVTec-AD

The MVTec Anomaly Detection (AD) dataset Bergmann et al. (2019a) constitutes a primary benchmark for unsupervised anomaly detection methods, particularly within industrial inspection contexts. It comprises of high-resolution images across 15 different objects, with pixel-wise annotations for a broad spectrum of defect types, making it invaluable for evaluating anomaly segmentation and classification algorithms. Contemporary methods evaluated on MVTec AD range from autoencoder-based reconstruction Yapp & Doan (2024); Cui et al. (2023), adversarial frameworks Akcay et al. (2018); Schlüter et al. (2022), and deep feature embeddings using pretrained vision models Roth et al. (2022); Zheng et al. (2022), to more recent approaches incorporating multiscale analysis and probabilistic scoring Ruff et al. (2018); Shin et al. (2023). Despite considerable advances, achieving generalizable anomaly detection performance across such diverse settings remains a central open challenge.

Continual learning in anomaly detection represents a rapidly evolving field, motivated by the need for robust systems that can adapt to evolving data distributions without catastrophic forgetting. In industrial and surveillance contexts, the nature of anomalies and normal patterns often changes over time, necessitating models capable of sequential adaptation while preserving learned knowledge from prior tasks.

3 METHODOLOGY

3.1 PROBLEM FORMULATION AND NOTATION

We study class-incremental anomaly detection on MVTec-AD, where each experience e_j introduces a single category, where only the data from that experience is accessible at training time. Training uses only defect-free ‘normal’ images X_{norm}^{train} of the current class, and evaluation mixes normal $\bigcup_{j=0}^{14} X_{norm, j=i}^{test}$ and anomalous images $X_{anomaly}^{test}$ for all seen classes with binary labels for anomaly detection.

Let $\mathcal{C}_t$ denote the set of classes observed up to and including experience t , and let $f_\theta(x, s, c)$ be a class-conditional diffusion model with de-noising step index $s \in \{0, \dots, S\}$ and class label $c \in \mathcal{C}_t$. The model is augmented with a Projector P that yields a whitened embedding $z(x) \in \mathbb{R}^d$ used for extreme value theory (EVT) calibration, where empirically $d = 2304$ due to the projection and spatial configuration at 32×32 resolution.

After each experience, per-class detectors are fit on distances $d_c(x) = \|z(x) - \mu_c\|_2^2$ computed over current-class embeddings, where μ_c is the class mean in the projected feature space. Test-time anomaly decisions use the model’s built-in classifier to predict $\hat{c}$ and then apply the corresponding per-class detector to produce cumulative probability scores which are then thresholded to produce binary anomaly predictions.

3.2 OUTLIER REJECTION USING WEIBULL MODELS

Following Mundt’s unified probabilistic framework for continual learning Mundt et al. (2022), we define acceptance regions in a shared latent space using EVT.

For each class c , a Weibull tail model is fit to the upper tail of $d_c(x)$ using libMR, inducing a per-class cumulative distribution:

$$s_c(d) = 1 - \exp\left(- (d/\lambda_c)^{k_c}\right)$$

This Weibull CDF score s_c is later thresholded to classify samples as ‘normal’ or ‘anomaly’.

3.3 GENERATION MEMORY BUDGET

We replace a persistent raw-image buffer with a regenerating, quality-gated generative buffer that is rebuilt after each experience. The system samples K images per previously seen class from the frozen teacher and filters them via the Weibull-based acceptance statistic (Weibull CDF $\leq \tau_{\text{gen}} = 0.5$) to obtain a replay set of effective size $M = K \cdot |\mathcal{C}_{j-1}|$ interleaved during current-task optimization, where $K = 100$.

3.4 MODEL ARCHITECTURE: DISTILLATION, DIFFUSION AND WEIBULL MODELS

As we can see in Figure 1, the backbone is a UNet2D with a Projector P and a linear classifier head H, enabling joint diffusion training and class prediction for EVT fitting in a shared feature space at 32×32 resolution with Adam learning rate 2×10^{-4} .

We save the teacher model after each experience and use DDIM sampling for replay generation. The student model then performs knowledge distillation from the teacher on the replayed samples.

At each experience, we expand class embeddings as needed, construct a Weibull-gated generative replay set from the frozen teacher, train the student on interleaved current and replay data with denoising and classification losses while keeping the projection frozen, fit the current-class Weibull, and evaluate across all seen experiences.

3.5 TRAINING ALGORITHM

Algorithm 1 Generative distillation with outlier rejection (Weibull)

Initialize: U-Net, Projector, Classifier for $\mathcal{C}_0 = \{c_0\}$ { Setup model for first class}
 Train on $X_{\text{norm},j=0}^{\text{train}}$, freeze Projector {Learn initial representations and fix the Projector P}
 Fit Weibull $W_1(\tau_0, \kappa_0, \lambda_0)$ on projected embeddings {Learn parameters for the distribution}
for $t = 1, \dots, 14$ **do**
 $\mathcal{C}_j \leftarrow \mathcal{C}_{j-1} \cup \{c_j\}$ {Add new class to set}
 Teacher $\leftarrow$ U-Net {Save current model}
 Expand U-Net, Classifier Head H to support $|\mathcal{C}_j|$ classes {Accommodate new class}
 Generate replay: $\mathcal{R}_t \leftarrow \emptyset$ {Initialize replay buffer}
 for $c \in \mathcal{C}_{j-1}$ **do**
 repeat
 $\tilde{x} \sim \text{DDIM}(\text{Teacher}, c)$ {Sample from frozen teacher}
 $z = \text{Projector}(\text{Teacher}(\tilde{x}))$ {Extract embedding}
 $s_c = W_c^{-1}(\|z - \mu_c\|_2)$ {Compute Weibull CDF score}
 until $s_c \leq \tau_{\text{gen}}$ and $|\mathcal{R}_j^{(c)}| = K$ {Accept if quality threshold met}
 $\mathcal{R}_j \leftarrow \mathcal{R}_j \cup \mathcal{R}_j^{(c)}$ {Add to replay set}
 end for
 Train: U-Net, Classifier on $X_{\text{norm},t}^{\text{train}} \cup \mathcal{R}_j$ {Joint training on current + replay}
 Minimize $\mathcal{L}_{\text{total}}(\theta)$
 Keep Projector frozen {Preserve embedding geometry}
 Fit EVT: $\mu_j \leftarrow \mathbb{E}[\text{Projector}(\text{U-Net}(X_{\text{norm},j}^{\text{train}}))]$ {Compute class mean}
 $d_j \leftarrow \{\|\text{Projector}(\text{UNet}(x)) - \mu_t\|_2^2 : x \in X_{\text{norm},j}^{\text{train}}\}$ {Distance to mean}
 $W_j \leftarrow \text{FitWeibull}(d_j, \text{tailsize})$
end for

4 EXPERIMENTAL SETUP

4.1 DATASET AND PROTOCOL

We evaluate on MVTec-AD Bergmann et al. (2019b) with 15 underlying classes making up the different objects represented in the data. Therefore, we adopt a class-incremental learning setup Masana et al. (2022) with 15 experiences, where each experience $e \in \{1, \dots, 15\}$ introduces exactly one class and contains only normal images from that category for training. The MVTec class here refers to the different categories of industrial objects in the MVTec-AD dataset and the anomaly class refers to the image being classified as 'normal' or 'anomaly' for the anomaly detection task which is what most of the results capture.

4.2 DATA SPLITS PER CATEGORY

- Train (normal-only): Official MVTec-AD training split
- Test (normal + anomaly): Official test split uses normal and anomalous images

We do not use any anomalous images for training, fitting or classification.

4.3 BACKBONE

We use a diffusion U-Net with 1000 training steps. When running the U-Net, we randomly initialize a timestep t and noise that will be added to the image. These are then passed to the U-Net which outputs the predicted noise which must be removed from the noisy image to de-noise it. We also use the U-Net for embeddings extraction which has been discussed later.

4.4 CLASSIFIER HEAD AND PROJECTED EMBEDDING SPACE

We attach a Projector P at the U-Net bottleneck featuring a 1×1 Conv2D Layer ($256 \rightarrow 64$ channels) and flattening. P is only trained in the first experience and later frozen to define a canonical coordinate system. All embeddings below are taken after P . The model is aided by a linear classifier head H which classifies samples into MVTec classes.

4.5 WHITENING

We normalize every projected embedding P by taking mean and variance over P and apply whitening before any distance operations are applied: $\hat{z} := \frac{(z - \mu)}{\sigma}$.

The Projector P coupled with Whitening makes sure that we map all embeddings to a shared embedding space. Additional losses have been added to ensure that this latent space stays fixed across experiences.

4.6 WEIBULL FITTING (EVT)

For each underlying MVTec class, c , from the inlier distances, we take the top 10% right-tail and fit a 2-parameter Weibull (k_c, λ_c) by MLE. All fitting is done in whitened space. At inference, we classify the sample using the classifier head H , pick the corresponding weibull model which had been fit during training and obtain the cumulative probability of the sample.

4.7 GENERATIVE REPLAY AND OUTLIER REJECTION

During training, we use generative replay to mitigate forgetting. At experience $e > 0$, the teacher (as a snapshot from $e - 1$) generates candidate clean x_0 samples conditioned on past classes. Each candidate is screened to have cumulative outlier probability less than a quality threshold τ in the projected whitened space. For the Weibull distribution corresponding to the class, it is accepted if it is below the Weibull quality threshold τ and rejected otherwise. Only accepted generations enter the replay buffer and are then used for generative distillation.

4.8 TRAINING LOSSES

- Diffusion Loss: Standard MSE between added noise (ϵ) and predicted noise ($\epsilon_\theta(\mathbf{x}_t, t)$) at a random t

$$\mathcal{L}_{\text{diff}}(\theta) = \mathbb{E}_{t, \mathbf{x}_0, \epsilon} [\|\epsilon - \epsilon_\theta(\mathbf{x}_t, t)\|_2^2] \quad (1)$$

- Classifier Loss: Cross-entropy loss for classification

$$\mathcal{L}_{\text{cls}}(\theta) = - \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\text{softmax}(f_\theta(\mathbf{h}_i))_c) \quad (2)$$

- Teacher Knowledge Distillation: MSE Between student and teacher noise predictions at the same (x_t, t)

$$\mathcal{L}_{KD}(\theta) = \mathbb{E}_{t, \mathbf{x}_0, \epsilon} [\|\epsilon_{\theta_s}(\mathbf{x}_t, t) - \epsilon_{\theta_t}(\mathbf{x}_t, t)\|_2^2] \quad (3)$$

- Embedding Alignment Loss - L2 norm between student and teacher projected embeddings

$$\mathcal{L}_{\text{align}}(\theta) = \mathbb{E}_{t, \mathbf{x}_0, \epsilon} [\|\mathbf{z}_{\theta_s}(\mathbf{x}_t, t) - \mathbf{z}_{\theta_t}(\mathbf{x}_t, t)\|_2^2] \quad (4)$$

- Center Loss - L2 norm between projected embedding and batch mean to ensure that the embeddings of a given class map to the similar vicinity in the embedding space

$$\mathcal{L}_{\text{center}}(\theta) = \frac{1}{N} \sum_{i=1}^N \|\mathbf{z}_i - \boldsymbol{\mu}_{\text{batch}}\|_2^2 \quad (5)$$

- Classifier Distillation Loss - KL divergence between teacher and student classifier head output classification probability distribution

$$\mathcal{L}_{\text{cl_distill}}(\theta) = \text{KL}(P_{\theta_t}(\mathbf{e}) \| P_{\theta_s}(\mathbf{e})) \quad (6)$$

$$\begin{aligned} \mathcal{L}_{\text{total}}(\theta) = & \mathcal{L}_{\text{diff}} + w_{\text{cls}} * \mathcal{L}_{\text{cls}} + w_{\text{center}} * \mathcal{L}_{\text{center}} \\ & + w_{\text{cl_distill}} * \mathcal{L}_{\text{cl_distill}} + w_{KD} * \mathcal{L}_{KD} \\ & + w_{\text{align}} * \mathcal{L}_{\text{align}} \end{aligned} \quad (7)$$

4.9 EVALUATION

We report per-experience and average Image-level AUPR, Accuracy and f1-score. Table 1 notes how these values vary across experiences. Table 2 compares our results with those from other baselines for this task.

4.10 ABLATIONS

- 4.1 Ablation on Training Epochs: We compare the 1 epoch results with the 100 epoch results
- 4.2 Ablation on Tailsize: We ablate on the percent of the train data that is considered as the tail while fitting the Weibull Distribution. We compare having 5% of the data as tail results with the 10% results

5 RESULTS

Main Results

Table 1: Performance metrics across continual learning experiences

Experience	AUPR	Accuracy	F1 Score
0	0.9673	0.7108	0.7736
1	0.7088	0.5322	0.5477
2	0.7233	0.7233	0.8394
3	0.7324	0.7324	0.8455
4	0.7996	0.6143	0.7333
5	0.7164	0.7164	0.8348
6	0.7204	0.7204	0.8375
7	0.7316	0.7316	0.8450
8	0.7491	0.7491	0.8565
9	0.7473	0.7395	0.8497
10	0.7458	0.7458	0.8544
11	0.7465	0.7391	0.8485
12	0.7213	0.7217	0.8384
13	0.7236	0.7236	0.8397
14	0.7293	0.7293	0.8434

Comparison with baselines

Table 2: F1-score comparison across anomaly detection models. **Best** results are bolded and worst results are underlined

Metric	DRÆM	PatchCore	PaDiM	CFA	STPFM	EfficientAD	FastFlow	Diff-Cont (Our)
F1-Score	0.87	0.97	<u>0.83</u>	0.94	0.89	0.89	0.89	0.84

Ablation Studies: On Epochs

Table 3: Performance metrics across training epochs

Epochs	AUPR	Accuracy	F1 Score
1	0.7293	0.7293	0.8434
100	0.7481	0.6928	0.8066

5.4 Ablation Studies: On Tailsize

Table 4: Performance metrics at different training completion percentages

Tailsize %	AUPR	Accuracy	F1 Score
5%	0.724	0.682	0.807
10%	0.7293	0.7293	0.8434

6 DISCUSSION

6.1 TRADEOFFS AND LIMITATIONS

Accuracy and f1-score can be seen to follow a similar trend wherein the scores gradually increase while the AUPR can be seen to decrease and flat-line around 70-75%. Our method exhibits competitive performance against baselines while maintaining a light-weight pipeline for industrial applications. This proves that the model is able to get better over time, while not catastrophically forgetting previous classes, which is crucial in the industry where data flows in large volumes. Our ablations show that training the fitting of the Weibull with a slightly higher tailsize percentage helps capture information about the distribution more accurately.

Despite our best efforts, there still appears to be room for improvement in our method, and we hypothesise that this could be achieved through ensuring the representations that the UNet learns for

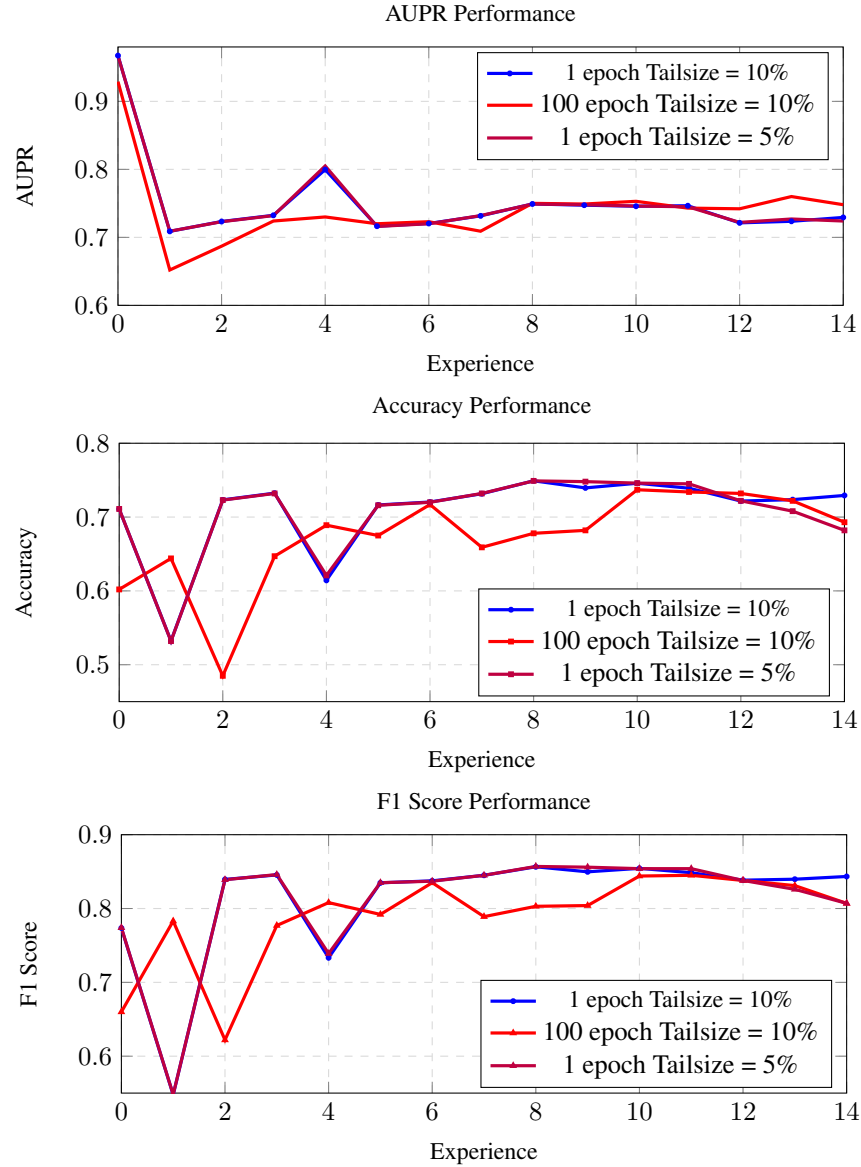


Figure 2: Variation of Performance Metrics across experiences

generative replay are of sufficiently high quality such that the replayed samples can be used in future experiences to tune the rejection thresholds for the per-class Weibull’s. Further to this, we believe that the projection space that the model uses could be improved to ensure rotational-invariance and meaningful embeddings which naturally distinguish between normal and anomalous samples on a per-class basis. Without proper representations with the correct amount of overlap between classes, the embedding space becomes either a lookup table (with too little overlap) or loses all meaning altogether (with too much). This could be

Why we chose U-Net bottleneck embeddings During the forward pass, we pass the noised images along with the added noise to the model. As a result, the bottleneck embedding can be seen not only as a representation of the image but also of the reconstruction error associated with the sample. We would expect the U-Net to take more epochs to train so that it can generate better quality images. However, because the Weibull distribution is fit on the projected embeddings, which capture reconstruction loss, we are able to achieve superior performance with less training.

7 CONCLUSION

We propose a light-weight, easy-to-train anomaly detection workflow for industrial applications using a combination of generative distillation, Diffusion-based methods and outlier rejection through Weibull models. On the MVTec-AD dataset, we demonstrate how our method achieves competitive results with baselines when comparing their f1-scores. The use of deep generative models allows us to perform continual learning without storing a replay buffer which makes our method attractive for industrial applications. We are optimistic that this will inspire further research into the overlap between Continual Learning, Diffusion models and Extreme Value Theory techniques - as we believe this to be a mathematically grounded approach to anomaly detection, leveraging the positive aspects of each respective field.

DECLARATION ON GENERATIVE AI

During the preparation of this work, the author(s) used OpenAI ChatGPT-4o, ChatGPT-5 and Claude Sonnet 4.0 in order to: Grammar and spelling check, Paraphrase and reword. After using these tool(s)/service(s), the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication’s content.

REFERENCES

- Samet Akcay, Amir Atapour-Abarghouei, and Toby P. Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training, 2018. URL <https://arxiv.org/abs/1805.06725>.
- Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9584–9592, 2019a. doi: 10.1109/CVPR.2019.00982.
- Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad – a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019b.
- Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, 33:15920–15930, 2020.
- Yajie Cui, Zhaoxiang Liu, and Shiguo Lian. Patch-wise auto-encoder for visual anomaly detection. In *2023 IEEE International Conference on Image Processing (ICIP)*, pp. 870–874. IEEE, October 2023. doi: 10.1109/icip49359.2023.10222183. URL <http://dx.doi.org/10.1109/ICIP49359.2023.10222183>.
- Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

- James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.
- Marc Masana, Xialei Liu, Bartłomiej Twardowski, Mikel Menta, Andrew D Bagdanov, and Joost Van De Weijer. Class-incremental learning: survey and performance evaluation on image classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(5):5513–5533, 2022.
- Martin Mundt, Iuliia Plushch, Sagnik Majumder, and Visvanathan Ramesh. Open set recognition through deep neural network uncertainty: Does out-of-distribution detection require generative classifiers? In *Proceedings of the IEEE/CVF international conference on computer vision workshops*, pp. 0–0, 2019.
- Martin Mundt, Iuliia Plushch, Sagnik Majumder, Yongwon Hong, and Visvanathan Ramesh. Unified probabilistic deep continual learning through generative replay and open set recognition. *Journal of Imaging*, 8(4):93, 2022.
- Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.
- David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019.
- Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.
- Karsten Roth, Latha Pemula, Joaquin Zepeda, Bernhard Schölkopf, Thomas Brox, and Peter Gehler. Towards total recall in industrial anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14318–14328, 2022.
- Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 4393–4402. PMLR, 10–15 Jul 2018. URL <https://proceedings.mlr.press/v80/ruff18a.html>.
- Tim Salimans and Jonathan Ho. Progressive distillation for fast sampling of diffusion models. *arXiv preprint arXiv:2202.00512*, 2022.
- Hannah M. Schlüter, Jeremy Tan, Benjamin Hou, and Bernhard Kainz. Natural synthetic anomalies for self-supervised anomaly detection and localization, 2022. URL <https://arxiv.org/abs/2109.15222>.
- Hanul Shin, Jung Kwon Lee, Jaehong Kim, and Jiwon Kim. Continual learning with deep generative replay. *Advances in neural information processing systems*, 30, 2017.
- Woosang Shin, Jonghyeon Lee, Taehan Lee, Sangmoon Lee, and Jong Pil Yun. Anomaly detection using score-based perturbation resilience. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 23315–23325, 2023. doi: 10.1109/ICCV51070.2023.02136.
- Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020a.
- Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. *arXiv preprint arXiv:2011.13456*, 2020b.

- Edward K.Y. Yapp and Ngoc C.N. Doan. Anomaly detection on mvtec ad using vq-vae-2. *Procedia CIRP*, 130:1809–1814, 2024. ISSN 2212-8271. doi: <https://doi.org/10.1016/j.procir.2024.10.320>. URL <https://www.sciencedirect.com/science/article/pii/S2212827124014781>. 57th CIRP Conference on Manufacturing Systems 2024 (CMS 2024).
- Ye Zheng, Xiang Wang, Rui Deng, Tianpeng Bao, Rui Zhao, and Liwei Wu. Focus your distribution: Coarse-to-fine non-contrastive learning for anomaly detection and localization. In *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2022. doi: 10.1109/ICME52920.2022.9859925.