

MATCH: Flow Matching for Multi-View Anomaly Detection

Mathis Kruse , Melissa Schween , and Bodo Rosenhahn 

Institute for Information Processing, L3S, Leibniz University Hannover, Germany
`{kruse,schween,rosenhahn}@tnt.uni-hannover.de`

Abstract. Detecting anomalies in industrial objects is an important topic for increasing production efficiency. More complex objects often require the analysis of several view points, which has led to the field of multi-view anomaly detection. We present *MATCH*, the first multi-view anomaly detection method based on Flow Matching (FM). With the ODE formulation of Flow Matching, we can estimate likelihoods and thereby derive an anomaly score to detect anomalies in multi-view image data at object, image, and pixel-level. The architectural flexibility of FM models allows us to efficiently transform features of different spatial sizes to the normal distribution. We evaluate thoroughly on the already established Real-IAD data set and are also the first to provide a comprehensive evaluation of popular anomaly detection methods for the MANTA-Tiny data set. *MATCH* achieves state-of-the-art performance in both anomaly detection and segmentation, all while running on consumer-level hardware. By omitting the costly divergence term needed for likelihood estimation, we ensure that *MATCH* is usable in real-time production scenarios. Lastly, several ablation studies are conducted to validate the methodological choices. Code: <https://github.com/m-kruse98/MATCH>

Keywords: Anomaly Detection · Flow Matching · Multi-View Data

1 Introduction

Any production pipeline may inevitably produce some defective products, either through human error or machine failures. Therefore, detecting anomalies has always been a key task in increasing the efficiency and robustness of any industrial manufacturing line. However, anomalies can be very rare and are often impossible to predict beforehand, while there is a vast availability of defect-free data. This has led to the birth of semi-supervised anomaly detection (AD), where models learn from normal data only.

With the rise of deep learning, several popular AD benchmarks, along with well-performing methods, have been proposed [6, 83]. The bulk of AD research focuses on single-view images, where objects are usually rigidly aligned in the exact same spot in front of the camera. However, as objects get more complex, a single camera may no longer suffice to detect all possible anomalies. More sophisticated production lines and the ubiquitous usage of imaging systems have

caused multi-view anomaly detection to gain more traction [64, 80]. This lets the multi-view AD models also detect anomalies, that may not be covered by just a single camera. The new setting of multi-view AD has not only spawned a variety of new data sets [22, 64, 80] but also many new AD methods [43, 44, 49]. An example of such a detection task is visualized in Fig. 1, where anomalies are only present in some views.

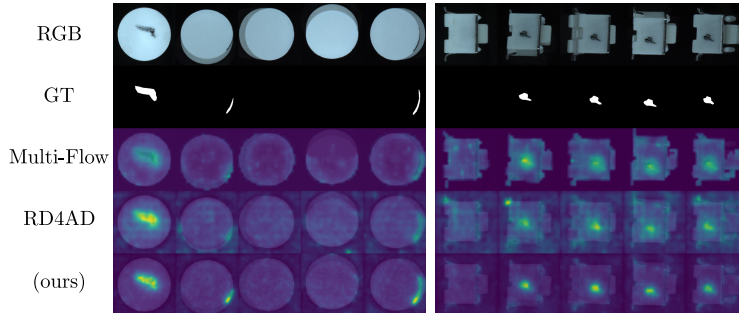


Fig. 1: Qualitative comparison of Multi-Flow [43], Reverse Distillation (RD4AD) [17], and *MATCH* (ours) on the *white tablet* and *short button* category of MANTA-Tiny. While Multi-Flow struggles with its foreground extraction and RD4AD detects some false positives, *MATCH* most accurately segments all anomalies.

Currently, a large variety of well-performing AD models use Normalizing Flows for estimating the likelihood of an image [27, 46, 57, 58, 68]. However, these Normalizing Flow models, which are almost always chosen to be models based on RealNVP [21], may struggle to handle the high dimensionality of (multi-view) images. Furthermore, their coupling block architecture limits them to a few transformation steps of expressivity. A new generative modelling technique termed *Flow Matching* [47] has begun to gain interest very quickly. Contrary to RealNVP, it can use more general network architectures and more easily represent even very high-dimensional data, while still functioning as both a generative model and a method for density estimation. This is especially interesting for the task of multi-view AD. Here, image data is usually very high dimensional, especially when different view points of an object need to be considered. In AD, very small deviations in the data may already constitute an anomaly. Thus, the decision boundary between normal and anomalous data may be a manifold of much smaller size, embedded within the high-dimensional data. This makes flow matching better suited for multi-view anomaly detection. A toy example, which visualizes the effect of learning a high-dimensional distribution, is shown in Fig. 2. Clearly, Flow Matching outperforms the inflexible RealNVP model, as it captures the underlying distribution even when it is embedded in higher dimensions.

Therefore, we propose **Multi-view Anomaly Detection using Flow Matching**, abbreviated to *MATCH*. To the best of our knowledge, it is the first Flow

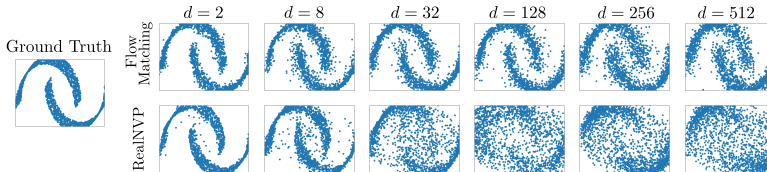


Fig. 2: Sampling the "Two Moons" data set, with Flow Matching and RealNVP. Similar to AD problems, where small deviances need to be detected in a high dimensional feature space, we embed the two-dimensional moons in d -dimensional space. While the expressiveness of Flow Matching handles high dimensions well, the less flexible RealNVP architecture starts to struggle early on. Experimental details are in the appendix.

Matching-based solution to multi-view AD. We leverage the capabilities in expressiveness to process feature maps of different spatial resolutions. Adding onto that, a few key simplifications are proposed to simplify the likelihood estimation and achieve new state-of-the-art results for both detecting and segmenting anomalies in popular multi-view AD benchmarks. All the while, our model still remains fast to use, reaching 18.77 FPS on consumer-level GPUs.

We summarize our contributions as follows:

- We propose a Flow Matching-based multi-view anomaly detection model, outperforming previous methods on both Real-IAD and MANTA-Tiny.
- We propose to omit the divergence term during likelihood estimation, which increases model throughput and thereby real-world usability, while leaving the detection performance unharmed and the score meaningful.
- We are the first to thoroughly benchmark the MANTA-Tiny data set.
- We release all code upon acceptance.

2 Related Work

We provide an overview of (multi-view) anomaly detection, as well as recent developments in Normalizing Flows, including Flow Matching.

2.1 Anomaly Detection

Overview. In recent years, research on anomaly detection (AD) has seen lots of success, with most of the popular benchmark data sets, such as MVTec AD [6] solved with almost perfect scores [4, 50, 56]. In almost all cases, AD is learned in a semi-supervised fashion, where only anomaly-free data is available during training, and one model is trained per object class. Some newer research started incorporating data modalities other than RGB, like depth information or normal maps [7, 11, 82]. Besides focusing on structural properties of single objects, some benchmarks also include logical anomalies, multiple objects, occlusions, and vast varieties of normal patterns [5, 36].

Besides data sets, there exist lots of categories of AD methods. Firstly, *Reconstruction-based* methods aim to reconstruct the image features through some bottleneck mechanism [8, 24, 29, 78]. Closely related are *student-teacher networks*, which distil information from a (potentially frozen) teacher network into a student network [4, 17, 59, 61, 77]. Some works also *incorporate anomalous information*, either real or synthetically generated, to cast the task into a supervised problem [19, 50, 70, 76]. As large vision-language-models (VLM) become more potent, *zero-shot AD* focuses on detecting anomalies without any prior knowledge [26, 39, 75, 81]. Using one model for all classes within a data set, a task termed *multi-class anomaly detection*, has also seen a rise in popularity [29, 31, 67].

Many methods, including ours, may be grouped into the field of *density estimation*. By learning the distribution of the normal training data, the likelihood of a sample is approximated and used as the anomaly score. Methods based on k-nearest Neighbors (kNN) or memory banks form one strand of research [14, 16, 45, 56], but rely heavily on the choice of features [35]. A more principled way of estimating the density is offered by Normalizing Flows (NF), which have seen wide application in AD [27, 43, 46, 57, 58, 68]. Here, the network explicitly learns the likelihood of a given sample using maximum likelihood estimation. A natural next step from NFs is Flow Matching [47], which is the core of our method as will be explained in Sections 2.2 and 3.

Multi-view Anomaly Detection. Mirroring real industrial scenarios with complex objects, a single camera may not suffice to capture all anomalies. Therefore, research on multi-view anomaly detection has started gaining traction, with the Real-IAD data set [64] serving as a new important benchmark. It shows each object from five different view points, with anomalies having to be detected both in an image-wise, as well as an aggregated object/sample-wise fashion. First methods for this exact setting have been proposed already [43], with some of them focused on multi-class AD [33].

Fan *et al.* present MANTA, a data set which features a camera setup very similar to Real-IAD [22]. MANTA covers a much higher variety of samples and categories, while also being magnitudes larger than anything before. Therefore, the authors also released MANTA-Tiny, a subset with largely the same test split, which is more easily usable for research purposes and solely used in this paper next to Real-IAD. To the best of our knowledge, we are the first to report comprehensive benchmark results on MANTA-Tiny with our experiments.

Another strand of research on multi-view AD uses more complex objects with dozens or up to hundreds of views per object [66, 79, 80]. Here, approaches based on novel-view synthesis methods have seen the most success [23, 44, 66, 80]. This remains, however, limited to one inflexible 3D model of normal samples, which fails when there is a larger variability of objects. This is something naturally addressed by most other anomaly detection methods, as they do not explicitly save one normal prototype to compare to. Concretely, density estimation techniques like Flow Matching allow for learning more complex distributions of normal samples.

There exist other variants of AD that intersect with multi-view/3D methods, such as finding one abnormal object out of many [10], leveraging 3D ground truth information [9], or dealing with complex lighting scenarios [12, 66]. All of these multi-view settings have significantly increased the difficulty in anomaly detection benchmarks and thereby inspired the research community.

2.2 Normalizing Flows

Preliminaries. The concept of Normalizing Flows [48] (NF) is based in the change of variables formula, which forms a relationship between a source distribution p (usually a standard Gaussian) and the target or data distribution q , which are both distributions over $\mathbb{R}^d$. The NF model $\phi : \mathbb{R}^d \rightarrow \mathbb{R}^d$ learns the mapping between data samples $x \sim q$ and p via

$$q(x) = p(\phi(x)) \left| \det \left[\frac{\partial \phi}{\partial x}(x) \right] \right|, \quad (1)$$

with $p(\cdot)$ and $q(\cdot)$ denoting the respective probability density functions (PDFs). Most notably, the flow ϕ , usually parametrized by a neural network, needs to be both invertible and have a Jacobian whose determinant is easy to compute. This has lead to the creation of specifically designed networks, that retain these properties [20, 41, 54, 55], with RealNVP being used the most in the anomaly detection literature for its ease of use [21, 43, 57].

Regular NFs usually chain together several invertible transformations $\phi = \phi_n \circ \dots \circ \phi_1$, limiting their expressivity by the number of chained flows n . This is alleviated with Continuous Normalizing Flows (CNF), which aim to learn the change in probability mass at every step using an ODE formulation [13]. The flow model now mimics the velocity field $u : \mathbb{R}^d \times [0, 1] \rightarrow \mathbb{R}^d$, learning the transformation needed to move a sample from p_0 at time $t = 0$ to $p_1 = q$ at time $t = 1$. In practice, however, CNFs are notoriously costly to train, requiring solving the entire ODE for a single training step [13, 25].

Flow Matching. With Flow Matching (FM) [47] training of CNFs has been simplified greatly by constructing conditional probability paths guiding from noise p to data q . Formally, an ODE is defined by its time-dependent vector field u_t , which is solved by its respective flow $\phi_t(x_0)$ as

$$\frac{d}{dt} \phi_t(x_0) = u_t(\phi_t(x_0)), \quad (2)$$

where the flow solution aims to satisfy

$$\phi_t(x_0) \sim p_t \text{ with initial condition } \phi_0(x_0) \sim p_0. \quad (3)$$

Thus, we aim to learn the vector field u_t , such that its flow ϕ_t generates the probability path from p_0 to p_1 , where each p_t is again a distribution over $\mathbb{R}^d$. As shown by Lipman *et al.* [47], the marginal vector field $u_t(x_0)$ turns out to

be equivalent to the expectation of the conditional vector field over the data distribution as

$$u_t(x_0) = \mathbb{E}_{x_1 \sim q} [u_t(x_0|x_1)]. \quad (4)$$

Thereby, sampling data $x_1 \sim q$ and regressing against the conditional vector field $u_t(x_0|x_1)$ suffices for learning a valid flow. This conditional vector field can easily be constructed given the available data, *e.g.* using Gaussian probability paths. This also results in the loss becoming a simple regression against the constructed conditional vector field. Subsequent work has explored different probability paths and other formulations, resulting in models with strong generative capabilities [15, 62].

3 Method

Problem Statement. During training, anomaly-free multi-view images of a certain object category are given. At inference, both anomalous and defect-free images appear, with the model having to detect the occurring anomalies. An object is considered anomalous if there is an anomaly visible in any of its views.

Objective. As motivated with Fig. 2, the distribution of normal images is learned using FM, specifically the optimal transport conditional Flow Matching (OT-CFM) formulation [62]. The loss for optimizing the parameters θ of the flow network u_t^θ is

$$\mathcal{L}_{\text{CFM}}^{\text{OT}} = \mathbb{E}_{\substack{x_0 \sim p_0, x_1 \sim q \\ t \sim \mathcal{U}[0,1]}} \|u_t^\theta(x_t) - (x_1 - x_0)\|^2, \quad (5)$$

where p_0 is standard normal Gaussian noise, q the data distribution and $x_t = (1-t)x_0 + tx_1$ an intermediate point along the flow.

Using the trained u_t^θ and any off-the-shelf ODE solver, we can use our algorithm in two ways. Running the ODE from $t = 0$ to $t = 1$ transforms noise $x_0 \sim p$ into data samples $x_1 \sim q$ (generative direction). Reversing it, and integrating the ODE from $t = 1$ to $t = 0$ lets us move real data samples to the Gaussian space, where we are able to measure likelihoods (density estimation direction). The likelihood of a sample x will serve as a proxy for the anomaly score and may be calculated as

$$\log p_1(x_1) = \log p_0(x_0) - \int_0^1 \text{div}(u_t^\theta)(x_t) dt, \quad (6)$$

where $p_0(x_0)$ is the density of the resulting sample in Gaussian space and $\text{div}(u_t^\theta)$ the divergence of the vector field at the intermediate solutions $x_t \sim p_t$ along which we integrate. In practice, this divergence term is estimated using methods such as Hutchinson’s trace estimator [38] or randomized quasi monte-carlo (RQMC) [53]. Even then, estimating the divergence term would result in (at the very least) one backpropagation per ODE step, which would vastly increase the computational requirements. This would leave the model unsuitable for deployment in real-time industrial production scenarios. An ablation conducted

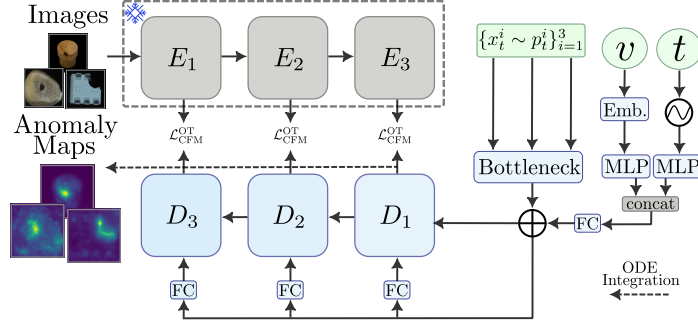


Fig. 3: Overview of *MATCH*. A frozen feature encoder E_i is used for latent embeddings. With these, intermediate latent vectors x_t^i are sampled, and, together with time t and view v information, fed through a bottleneck, and into the decoder layers D_i . The Flow Matching loss $\mathcal{L}_{\text{CFM}}^{\text{OT}}$ is applied and lets our model learn the trajectory from data space q to Gaussian space p . Running the model with an ODE from $t = 1$ to 0 lets us produce anomaly maps, which segment any anomalous regions.

in Sec. 4.4 shows that omitting the divergence term greatly saves computation time and memory while still performing well for AD.

Thus, our anomaly score just becomes $S(x) = -\log p_0(x_0)$, which in practice, calculated along the feature dimension, becomes

$$S(x_1) = -\log p_0(x_0) = \frac{\|x_0\|^2}{2} + \frac{d}{2} \log(2\pi), \quad (7)$$

where x_0 is the sample in Gaussian space, and d the feature dimensionality. This score may be interpreted, up to the additive constant, as the squared Mahalanobis distance of the sample x_0 to the standard Gaussian $\mathcal{N}(0, I)$. This interpretation is used in many related papers on AD and normalizing flows [27, 57, 58].

3.1 Model Architecture

Normalizing flows trained on raw images have been shown to fail in detecting anomalies [42]. Further, Flow Matching, just as diffusion models, performs well on latent feature distributions [15]. Therefore, the flow is instead learned on the latent distribution of a WideResNet [69] (pre-trained on ImageNet [18]) which serves as a frozen feature extractor.

We take inspiration by the *ReverseDistillation* (RD) architecture [17], with a high-level schematic of our adapted architecture shown in Fig. 3. A longer and completely in-depth graph of all layers and blocks can be found in the supplement in Sec. A. Firstly, the data samples are extracted from the first three layers E_i of the WideResNet as $\{x_1^i \sim q^i\}_{i=1}^3$, where i denotes the current layer of the extractor. To train, a time step $t \in \mathcal{U}(0, 1)$ and noise of the same size as the latent features $\{x_0^i \sim p^i\}_{i=1}^3$ are sampled. Then, the intermediate features $\{x_t^i = (1-t)x_0^i + tx_1^i\}_{i=1}^3$ are constructed as a linear combination and

fed through a bottleneck. After that, several ResNet-style layers D_i in reversed order are used to reconstruct features to the correct spatial dimension, similar to the architecture for RD4AD [17].

Time and View Embedding. To make the model usable for Flow Matching, we embed the time step information t using a 1024-dimensional sinusoidal positional encoding [63] and a simple two-layer MLP with SiLU activation [37]. Similarly, the index of the current view v is fed into a learned embedding layer as well as an MLP with the same dimensions. Both view and time information are concatenated, passed through another fully connected (FC) layer with SiLU activation and added to the output of the bottleneck. They also serve as input to the reconstruction layers D_i , with each layer having their own FC and SiLU layer for fine-tuning features to their needs.

Bottleneck Network. The input features $\{x_t^i \in \mathbb{R}^{C_i \times H_i \times W_i}\}_{i=1}^3$ are fed into the bottleneck network, consisting of an alignment network and bottleneck blocks. For the *alignment*, the map at $i = 3$ has the lowest feature resolution. We thereby run the other maps through a down sampling using 3×3 convolutional layers with stride 2, using one layer for $i = 2$, and two for $i = 1$. The spatially aligned maps are then concatenated to produce one feature vector. Each of the mentioned convolutional layer consist of its respective convolution followed by group normalization with 32 groups, and ReLU activation. Afterwards, three ResNet-style *bottleneck* blocks [17, 34] (again with group normalization) further distil the feature representation, forcing the network to focus on the normal structures found in training only. Residual connections are either directly implemented, or using a downsampling with another convolution.

Decoder Blocks. The decoder consists of three layers D_i each consisting of d_i decoder blocks, that progressively upsample the bottleneck feature representation back to the original spatial dimensions, which mostly follows RD [17, 72]. These decoder blocks use 1×1 convolutions followed by 3×3 convolutions with a stride of 2 for upsampling [72], group normalization, ReLU activation and a final 1×1 convolution. Here, the time-view embedding is again added to the latent features. Again, the residual connection may be done either using an identity map or a single transposed convolutional layer for upsampling.

Anomaly Scores. We produce three anomaly score maps $M_i \in \mathbb{R}^{H_i \times W_i}$ by calculating the anomaly score along the feature dimension of our decoded features as in Eq. (7). All of these maps are resized to the full spatial dimension of our images using bilinear interpolation. For segmentation, we simply use the combination of all maps $M_{\text{segment}} = \sum_i M_i$ as our final segmentation map. This includes both good segmentation of high-level defects, as well as the tiny low-level and hard-to-spot anomalies.

For detection, it has been shown that higher level representations are better suited to produce a single per-image anomaly score [42]. Thereby, we use $M_{\text{detect}} = M_2 + M_3$ as the sum of only the last two anomaly maps, which mostly focuses on higher-level concepts while not being influenced too much by lower level deviations. The final *image-wise* anomaly score is then produced by taking the maximum value found in this detection map M_{detect} . To produce a

sample-wise anomaly score, we again take a maximum across all anomaly scores produced for one multi-view instance.

4 Experiments

We evaluate our method on data sets Real-IAD [64] and MANTA-Tiny [22], while also reporting numbers for many competing state-of-the-art approaches.

4.1 Data Set and Metrics

We use both Real-IAD [64] and MANTA-Tiny [22] for evaluating our method. Real-IAD [64] is a large multi-view anomaly detection data set, made up of 30 different object classes taken from industrial production contexts, totalling roughly 100k normal and 50k anomalous images. Defects such as scratches, missing parts or deformations are some of the most common categories of anomalies.

MANTA [22] contains multi-view images of *tiny objects*, with the different object classes spanning across the categories of agriculture, medicine, electronics, mechanics, and groceries. With MANTA being about five times the size of Real-IAD, a smaller benchmark called MANTA-Tiny has been proposed, where images are resized and the amount of training instances is reduced to roughly match previous data sets and increase its popularity. The test sets of both splits remain largely the same. Depending on the category, defects may appear as misprints, stains, scratches, mold, or mildew, amongst many other possible faults.

For quantitative results, we report detection performance using the Area Under the Receiver Operating Curve (AUROC), which is independent of a specific threshold. The AUROC will be reported both *image-wise* (*i.e.* each image is treated independently, named I-AUROC) and *sample-wise*, where all images belonging to the same object are grouped together to form one single anomaly score (named S-AUROC). Lastly, for segmenting the anomalies, we also report the AUROC in a *pixel-wise* fashion (P-AUROC), as well as P-AP (average precision), and the more balanced P-AUPRO metric [6].

4.2 Implementation Details

All experiments are run on a Linux machine with a RTX 3090 consumer GPU. All metrics are calculated efficiently using the GPU-optimized ADEval library [74]. We resize all images to 256×256 pixels, resulting in extracted feature maps of sizes $(256 \times 64 \times 64)$, $(512 \times 32 \times 32)$, and $(1024 \times 16 \times 16)$ using the WideResNet50.

For optimization we use the AdamW optimizer [51] with a learning rate of $1e-4$, momentum parameters $\beta_{1,2} = (0.9, 0.95)$, and weight decay set to 0.01. Contrary to most NF models, we do not need any clipping or scaling for the gradients [2]. The decoder layers D_i consists of 3, 4 and 6 decoder blocks respectively. All convolutional layers work with a base width of 768 dimensions. We train on both data sets for 150 epochs with a batch size of 8. For inference using the ODE, we use the simple Euler method with step sizes of 0.2, resulting in 5

Table 1: Both segmentation and image-wise detection metrics averaged across all classes of Real-IAD [64]. The best performing method is marked in **bold** with the runner-up underlined. Our method is run $n = 5$ times.

Method	Detection Metrics ($\uparrow$)		Segmentation Metrics ($\uparrow$)		
	I-AUROC	S-AUROC	P-AUROC	P-AUPRO	P-AP
CFA [45]	71.44	77.53	95.58	79.05	13.63
DSR [71]	58.86	52.83	75.47	56.02	17.03
EfficientAD [4]	77.26	86.91	89.64	71.31	13.18
FastFlow [68]	79.26	92.30	90.18	66.99	06.16
Multi-Flow [43]	<u>90.27</u>	95.85	96.47	87.91	12.42
PaDiM [16]	89.27	91.33	94.22	75.33	06.73
RD4AD [17]	89.26	93.46	<u>98.81</u>	92.70	<u>32.65</u>
SimpleNet [50]	86.37	94.89	<u>93.90</u>	<u>78.87</u>	<u>18.36</u>
STFPM [65]	71.74	74.99	91.41	74.47	15.83
<i>MATCH(ours)</i>	91.17 ± 0.08	<u>95.63</u> ± 0.09	99.24 ± 0.01	94.76 ± 0.03	32.68 ± 0.03

forward passes per inference batch. We will later make this choice more founded in an ablation in Sec. 4.4.

Baselines. A variety of state-of-the-art AD methods are reproduced using the Anomalib [1] and ADer [74] benchmarking libraries. We chose them to be executed with the same consumer-level GPUs and setup as our method. The baselines include CFA [45], DSR [71], EfficientAD [4], FastFlow [68], Multi-Flow [43], PaDiM [16], ReverseDistillation (RD4AD) [17], SimpleNet [50], and STFPM [65].

4.3 Anomaly Detection Results

We run *MATCH* on both Real-IAD and MANTA-Tiny. For Real-IAD, all metrics are reported in Tab. 1. Here, we outperform all competitors in most metrics by convincing margins, reaching an I-AUROC of 91.17 for detection, and an P-AUPRO of 94.76 for segmentation, all while being very stable with a standard deviation of .08 and .03 respectively. Most notably, the direct flow-based competitor Multi-Flow, based on RealNVP, underperforms in detection by roughly one point, while *MATCH* is far more precise at segmentation, as visible in the P-AUPRO. This strengthens us in our claim that FM is better able to handle the high dimensionality of the feature maps, as it is still able to pinpoint the locations of the occurring anomalies.

As for the sample-wise detection (with S-AUROC), we equip all baselines with a simple max-aggregation, just as used by *MATCH*, to calculate their scores. We achieve an S-AUROC of 95.63 on Real-IAD. This score only lacks slightly behind the most direct competitor Multi-Flow, which is also optimized for multi-view AD, by .2 points. On all other metrics and tasks, we outperform the baselines, especially in segmentation tasks.

Segmentation results for the newly proposed and until now not thoroughly evaluated MANTA-Tiny data set are presented in Tabs. 2 and 3. The most im-

Table 2: Anomaly Segmentation results on MANTA-Tiny measured as **P-AUPRO** ($\uparrow$) and grouped by category. The best performing method is marked in **bold** with the runner-up underlined. Our method is run $n = 5$ times.

Category	CFA [45]	DSR [71]	Effic.AD [4]	FastFlow [68]	Multi-Flow [43]	PaDiM [16]	RD4AD [17]	SimpleNet [50]	STFPM [65]	<i>MATCH</i> (<i>ours</i>)
Agriculture	43.07	50.07	54.92	48.33	69.41	63.57	<u>69.74</u>	59.92	48.22	75.52 ± 0.05
Electronics	59.62	57.21	89.30	72.18	84.35	84.01	<u>92.79</u>	84.71	56.91	93.93 ± 0.02
Groceries	39.35	46.43	<u>66.72</u>	43.81	63.28	55.23	56.52	61.78	34.48	70.03 ± 0.05
Mechanics	71.35	65.34	87.72	76.98	87.39	86.98	<u>94.41</u>	85.47	80.47	96.35 ± 0.02
Medicine	67.93	58.46	78.26	75.66	<u>83.38</u>	79.71	82.75	80.11	70.20	89.10 ± 0.04
MANTA-Tiny	61.64	58.26	80.83	69.65	81.76	79.43	<u>85.59</u>	79.42	64.19	89.66 ± 0.02

Table 3: Anomaly Segmentation results on MANTA-Tiny measured as **Pixel-AUROC** ($\uparrow$) and grouped by category. The best performing method is marked in **bold** with the runner-up underlined. Our method is run $n = 5$ times.

Category	CFA [45]	DSR [71]	Effic.AD [4]	FastFlow [68]	Multi-Flow [43]	PaDiM [16]	RD4AD [17]	SimpleNet [50]	STFPM [65]	<i>MATCH</i> (<i>ours</i>)
Agriculture	70.84	63.73	84.50	71.67	84.92	83.36	<u>87.22</u>	87.19	76.60	87.50 ± 0.02
Electronics	86.60	76.05	95.63	91.39	95.28	95.62	<u>98.28</u>	96.49	83.18	98.47 ± 0.00
Groceries	73.48	64.92	86.78	71.05	85.58	75.25	83.26	<u>86.76</u>	67.36	85.92 ± 0.04
Mechanics	87.09	70.32	94.73	91.59	97.00	95.9	<u>98.28</u>	96.41	92.70	98.55 ± 0.00
Medicine	82.85	67.54	90.17	89.62	91.70	91.62	<u>92.66</u>	91.87	89.54	95.53 ± 0.02
MANTA-Tiny	83.16	70.20	92.21	87.35	93.07	91.85	<u>94.60</u>	93.63	85.50	95.65 ± 0.01

portant results on segmentation are presented with the P-AUPRO in Tab. 2. Here, we achieve a P-AUPRO of 89.66 across all categories, which outperforms all baselines. We further note that classes containing a larger variance in their normal samples, such as the naturally produced agriculture and groceries categories, are far harder to solve than categories focusing on classical industrial contexts. Also, the ReverseDistillation (RD4AD) model almost always comes in second, highlighting the effectiveness of the chosen architecture. Still, the Flow Matching formulation managed to gain very strong improvements, with a P-AUPRO more than four points higher.

Only in the detection metric I-AUROC does *MATCH* fall behind, as SimpleNet is able to get a score of 93.19 while *MATCH* reaches a competitive 90.66. We posit that this may be due to the much higher variance in defects found in MANTA-Tiny, as some tiny defects may be harder to translate into a high score for the entire image. Handling this issue may be interesting for future work. Lastly, we also evaluated our model (trained on MANTA-Tiny) on the full MANTA test split, which is roughly 70% identical. We observed no noticeable performance drops, as shown in the supplementary. This shows MANTA-Tiny to be a good real-world benchmark, that is worth exploring.

For both of the mentioned benchmarks, we provide a variety of qualitative anomaly maps in Fig. 4. Here, the high segmentation quality of *MATCH* becomes apparent, which is a clear advantage to preceding Normalizing Flow-based methods [43, 58]. We also provide additional qualitative segmentations for every

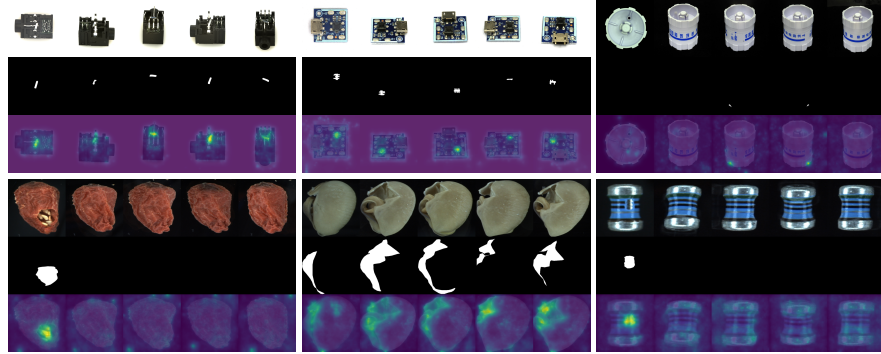


Fig. 4: A selection of qualitative anomaly maps for classes audiojack, PCB, and regulator of Real-IAD (top row). The bottom row shows goji berries, soybeans, and wafer resistors of MANTA-Tiny. Pictured are the original RGB image, ground truth segmentation maps and the resulting anomaly maps overlaid on RGB.

category of MANTA-Tiny, as well as an analysis of some failure cases in the supplement in Sec. F, together with the comprehensive results for all metrics on both data sets in Sec. G.

4.4 Ablation Studies

We conduct ablations to motivate design choices and provide further insight. Fig. 5 shows the impact of different base widths of our network. The AD performance being best around a width of 768 confirms that strong and expressive (FM-based) models are most suited. Also, matching the parameter count of RD4AD to *MATCH*, RD4AD reaches an I-AUROC of 86.75, confirming that gains stem from the FM formulation, and not just model capacity. We do not consider higher widths, as we want *MATCH* to remain usable on affordable consumer-level hardware.

Impact of the Divergence Term. As described in Sec. 3, we omit the divergence term in Eq. (6), which is equal to the trace of the Jacobian matrix $\text{tr}[\partial_x u_t(x)]$, to increase the computational efficiency of our model. There are several possible methods to estimate this divergence term. Using Hutchinson’s trace estimator, we may sample M vectors $z \sim \mathcal{Z} = \mathcal{N}(0, I)$ and calculate

$$\text{div}(u_t)(x) = \text{tr}[\partial_x u_t(x)] = \mathbb{E}_{z \sim \mathcal{Z}} \text{tr}[z^T \partial_x u_t(x) z] \quad (8)$$

as an unbiased estimator of the divergence [48], which is added to the anomaly score to form the likelihood. By increasing the number of random vectors $M > 0$, we get a more accurate estimation.

While Hutchinson’s estimator draws its z vectors independent of each other, we may try to decrease the variance in our prediction by sampling them from

a dependent distribution. This is the basis for randomized quasi monte-carlo (RQMC) [53]. For the vectors for RQMC we use Latin hypercube sampling [52], due to the high dimensionality of our feature vectors. The results for all estimation methods are shown in Tab. 4 for different values for M .

Including the divergence does not lead to any noticeable gains for the detection AUROC. The same behavior applies to all metrics, as completely shown in the supplementary materials. Most interestingly, with $M = 1$, the divergence only makes up a total of .57% of the final score magnitude

on Real-IAD, since $\frac{|\text{div}(u_t^\theta)|}{|\log(x_0)|} = 0.0057$. Moreover, leaving it out gets a boost of at least three times more model executions per second. Also, increasing M linearly increases both execution time and memory requirements, while also not gaining any precision in detection. We hypothesize that our model learns a very smooth vector field flowing from p_0 to p_1 , which has almost no noticeable local divergence. Furthermore, this strengthens us in our decision to leave RealNVP-style models behind, which rely on architectures specifically designed to make this divergence tractable. Lastly, the flexibility gained with Flow Matching (as was visualized in Fig. 2) also provides an advantage on important real-world benchmarks, when comparing it to its more constrained RealNVP competitors.

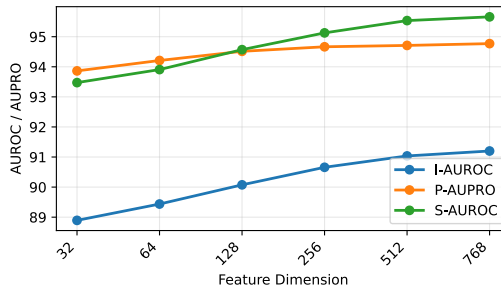


Fig. 5: Varying the feature dimension for bottleneck and decoder from 32 to 768. All of the metrics linearly rise with higher dimensions.

Impact of ODE Solver. We test a variety of popular methods for solving ODEs on the Real-IAD data set. For this we choose methods with fixed step sizes, namely the Euler method, the Midpoint method, and the fourth order variant of Runge-Kutta (RK4) [30]. We vary their step sizes from 0.2 to 0.05 and record their impact on AD performance and model throughput (as FPS) in Tab. 5. We also compare to the adaptive step version of Runge-Kutta (Dopri5). Additional adaptive solvers, such as Euler-Heun, Fehlberg2, or Bosh3 achieved similar scores, but are reported in the supplementary for brevity.

Going from Euler to Midpoint to RK4, the amount of function evaluations per batch doubles. This is also visible in the throughput, which gives the Euler method a clear edge for fast in-production use, since it is the only method capable of processing more than 10 images per second. The adaptive Dopri5-solver is slower than all competing methods while not bringing any more precision to the detection. However, none of the methods show an all-too-clear advantage in the measured AUROC except for the Euler method. Here, a small improvement can be measured only for the higher step sizes. All other methods achieve an AUROC of roughly 91.06, irrespective of step size.

Table 4: I-AUROC for the divergence term from Eq. (6) on Real-IAD, with varying number of samples M . Without the divergence, FPS & average memory improves, while detection is unchanged.

Method	$M = 0$	$M = 1$	$M = 2$	$M = 4$
<i>MATCH</i>	91.172	-	-	-
Hutchinson [38]	-	91.170	91.169	91.170
RQMC [53]	-	91.170	91.170	91.170
($\uparrow$) FPS	18.77	5.40	3.16	1.56
($\downarrow$) Memory (GB)	8.88	10.28	10.70	11.45

Table 5: I-AUROC for different ODE solvers with varying step size parameter on Real-IAD. Increasing the step size also linearly decreases the amount samples processed per second.

Metric	I-AUROC ($\uparrow$)			FPS ($\uparrow$)		
Step size	0.05	0.1	0.2	0.05	0.1	0.2
Euler	91.09	<u>91.12</u>	91.17	3.26	<u>6.54</u>	13.13
Midpoint	91.06	91.06	91.08	1.63	3.27	6.55
RK4	91.06	91.07	91.07	0.62	1.24	2.48
Dopri5	$\leftarrow$ 91.06 $\rightarrow$			$\leftarrow$ 0.48 $\rightarrow$		

Clear hypotheses can, however, not be made, as the AUROC improvement with Euler lies roughly within the standard deviation reported in Tab. 1. Nevertheless, the simplicity still makes the Euler method the fastest and best usable ODE solver for our AD use case, as there is no trade-off between speed and accuracy. It is also consistent with previous research, which showed no substantial numerical errors, when using simpler ODE solvers [47].

Supplementary Experiments. We conduct three additional experiments in our *supplementary* materials in Secs. B to D, to further validate our model’s strength. Firstly, we report metrics for the classic MVTec AD benchmark in the appendix, where *MATCH* is still competitive. Furthermore, we also test the popular multi-class AD setting. Even though *MATCH* is not explicitly focused on multi-class AD, it achieves a very competitive I-AUROC of 87.5 on Real-IAD, only slightly beaten by Dinomaly [29]. Lastly, we highlight the robustness to other backbones, by swapping *MATCH* and PaDiM & RD4AD to a DINOv3 [60] backbone. Even then, the AD performance stays the same or deteriorates slightly. Furthermore, with all methods fixed to a DINOv3 backbone, *MATCH* still outperforms the other methods. Still, we prefer the classic WideResNet due to its lower computational costs.

5 Conclusion

We have proposed *MATCH*, an effective framework for detecting anomalies in the multi-view setting. Using the simple Flow Matching formulation and a multi-layer feature extraction pipeline, we are able to establish state-of-the-art results on Real-IAD, as well as new comprehensive benchmarking results on MANTA-Tiny. Future work will focus on including more prior knowledge, such as explicit 3D constraints or multi-view correspondences, into the Flow Matching pipeline. More recent improvements in Flow Matching and generative modelling, such as energy matching [3], may also suit themselves to future endeavours in anomaly detection.

Acknowledgements

This work was supported by the MWK of Lower Saxony within Hybrint (VWZN4219) and LCIS (VWZN4704), the Deutsche Forschungsgemeinschaft (DFG) under Germany’s Excellence Strategy within the Cluster of Excellence PhoenixD (EXC2122) and Quantum Frontiers (EXC2123), the European Union under grant agreement no. 101136006 – XTREME.

References

1. Akcay, S., Ameln, D., Vaidya, A., Lakshmanan, B., Ahuja, N., Genc, U.: Anomali: A deep learning library for anomaly detection. In: 2022 IEEE International Conference on Image Processing (ICIP). pp. 1706–1710. IEEE (2022)
2. Ardizzone, L., Lüth, C., Kruse, J., Rother, C., Köthe, U.: Guided image generation with conditional invertible neural networks. arXiv preprint arXiv:1907.02392 (2019)
3. Balcerak, M., Amiranashvili, T., Terpin, A., Shit, S., Bogensperger, L., Kaltenbach, S., Koumoutsakos, P., Menze, B.: Energy matching: Unifying flow matching and energy-based models for generative modeling. arXiv preprint arXiv:2504.10612 (2025)
4. Batzner, K., Heckler, L., König, R.: Efficientad: Accurate visual anomaly detection at millisecond-level latencies. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 128–138 (January 2024)
5. Bergmann, P., Batzner, K., Fauser, M., Sattlegger, D., Steger, C.: Beyond dents and scratches: Logical constraints in unsupervised anomaly detection and localization. *International Journal of Computer Vision* **130**(4), 947–969 (2022)
6. Bergmann, P., Fauser, M., Sattlegger, D., Steger, C.: Mvtec ad—a comprehensive real-world dataset for unsupervised anomaly detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 9592–9600 (2019)
7. Bergmann, P., Jin, X., Sattlegger, D., Steger, C.: The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization. In: Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2022, Volume 5: VISAPP, Online Streaming, February 6-8, 2022. pp. 202–213. SCITEPRESS (2022)
8. Bergmann, P., Löwe, S., Fauser, M., Sattlegger, D., Steger, C.: Improving unsupervised defect segmentation by applying structural similarity to autoencoders. In: Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2019, Volume 5: VISAPP, Prague, Czech Republic, February 25-27, 2019. pp. 372–380 (2019)
9. Bhunia, A., Li, C., Bilen, H.: Looking 3d: Anomaly detection with 2d-3d alignment. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024. pp. 17263–17272. IEEE (2024). <https://doi.org/10.1109/CVPR52733.2024.01634>, <https://doi.org/10.1109/CVPR52733.2024.01634>
10. Bhunia, A., Li, C., Bilen, H.: Odd-one-out: Anomaly detection by comparing with neighbors. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025. pp.

- 20395–20404. Computer Vision Foundation / IEEE (2025). <https://doi.org/10.1109/CVPR52734.2025.01899>, https://openaccess.thecvf.com/content/CVPR2025/html/Bhunias_Odd-One-Out_Anomaly_Detection_by_Comparing_with_Neighbors_CVPR_2025_paper.html
11. Bonfiglioli, L., Toschi, M., Silvestri, D., Fioraio, N., De Gregorio, D.: The eyecandies dataset for unsupervised multimodal anomaly detection and localization. In: Proceedings of the 16th Asian Conference on Computer Vision (ACCV2022 (2022)), aCCV
 12. Cao, Y., Cheng, Y., Xu, X., Zhang, Y., Sun, Y., Tan, Y., Zhang, Y., Huang, X., Shen, W.: Visual anomaly detection under complex view-illumination interplay: A large-scale benchmark. arXiv preprint arXiv:2505.10996 (2025)
 13. Chen, R.T., Rubanova, Y., Bettencourt, J., Duvenaud, D.K.: Neural ordinary differential equations. *Advances in neural information processing systems* **31** (2018)
 14. Cohen, N., Hoshen, Y.: Sub-image anomaly detection with deep pyramid correspondences. arXiv preprint arXiv:2005.02357 (2020)
 15. Dao, Q., Phung, H., Nguyen, B., Tran, A.: Flow matching in latent space. arXiv preprint arXiv:2307.08698 (2023)
 16. Defard, T., Setkov, A., Loesch, A., Audigier, R.: Padim: a patch distribution modeling framework for anomaly detection and localization. In: International Conference on Pattern Recognition. pp. 475–489. Springer (2021)
 17. Deng, H., Li, X.: Anomaly detection via reverse distillation from one-class embedding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9737–9746 (2022)
 18. Deng, J., Dong, W., Socher, R., Li, L.J., Li, K., Fei-Fei, L.: Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. pp. 248–255. Ieee (2009)
 19. Ding, C., Pang, G., Shen, C.: Catching both gray and black swans: Open-set supervised anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18–24, 2022. pp. 7378–7388. IEEE (2022). <https://doi.org/10.1109/CVPR52688.2022.00724>, <https://doi.org/10.1109/CVPR52688.2022.00724>
 20. Dinh, L., Krueger, D., Bengio, Y.: NICE: non-linear independent components estimation. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7–9, 2015, Workshop Track Proceedings (2015), <http://arxiv.org/abs/1410.8516>
 21. Dinh, L., Sohl-Dickstein, J., Bengio, S.: Density estimation using real NVP. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24–26, 2017, Conference Track Proceedings. OpenReview.net (2017), <https://openreview.net/forum?id=HkpbH91x>
 22. Fan, L., Fan, D., Hu, Z., Ding, Y., Di, D., Yi, K., Pagnucco, M., Song, Y.: MANTA: A large-scale multi-view and visual-text anomaly detection dataset for tiny objects. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11–15, 2025. pp. 25518–25527. Computer Vision Foundation / IEEE (2025). <https://doi.org/10.1109/CVPR52734.2025.02376>
 23. Galappaththige, C.J., Lai, J., Windrim, L., Dansereau, D.G., Sünderhauf, N., Miller, D.: Multi-view pose-agnostic change localization with zero labels. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11–15, 2025. pp. 11600–11610. Computer Vision Foundation / IEEE (2025). <https://doi.org/10.1109/CVPR52734.2025.01083>, https://openaccess.thecvf.com/content/CVPR2025/html/Galappaththige_

- Multi-View-Pose-Agnostic-Change-Localization-with-Zero-Labels-CVPR-2025_paper.html
24. Gong, D., Liu, L., Le, V., Saha, B., Mansour, M.R., Venkatesh, S., van den Hengel, A.: Memorizing normality to detect anomaly: Memory-augmented deep autoencoder for unsupervised anomaly detection. In: 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019, Seoul, Korea (South), October 27 - November 2, 2019. pp. 1705–1714. IEEE (2019). <https://doi.org/10.1109/ICCV.2019.00179>, <https://doi.org/10.1109/ICCV.2019.00179>
 25. Grathwohl, W., Chen, R.T.Q., Bettencourt, J., Sutskever, I., Duvenaud, D.: FFJORD: free-form continuous dynamics for scalable reversible generative models. In: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net (2019), <https://openreview.net/forum?id=rJxgknCcK7>
 26. Gu, Z., Zhu, B., Zhu, G., Chen, Y., Tang, M., Wang, J.: Anomalygpt: Detecting industrial anomalies using large vision-language models. In: Wooldridge, M.J., Dy, J.G., Natarajan, S. (eds.) Thirty-Eighth AAAI Conference on Artificial Intelligence, AAAI 2024. pp. 1932–1940. AAAI Press (2024). <https://doi.org/10.1609/AAAI.V38I3.27963>, <https://doi.org/10.1609/aaai.v38i3.27963>
 27. Gudovskiy, D., Ishizaka, S., Kozuka, K.: Cflow-ad: Real-time unsupervised anomaly detection with localization via conditional normalizing flows. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 98–107 (2022)
 28. Guo, J., Lu, S., Jia, L., Zhang, W., Li, H.: Recontrast: Domain-specific anomaly detection via contrastive reconstruction. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023 (2023), http://papers.nips.cc/paper_files/paper/2023/hash/228b9279ecf9bbafe582406850c57115-Abstract-Conference.html
 29. Guo, J., Lu, S., Zhang, W., Chen, F., Li, H., Liao, H.: Dinomaly: The less is more philosophy in multi-class unsupervised anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025. pp. 20405–20415. Computer Vision Foundation / IEEE (2025). <https://doi.org/10.1109/CVPR52734.2025.01900>, https://openaccess.thecvf.com/content/CVPR2025/html/Guo_Dinomaly_The_Less_Is_More_Philosophy_in_Multi-Class_Unsupervised_Anomaly_CVPR_2025_paper.html
 30. Hairer, E., Wanner, G., Nørsett, S.P.: Solving ordinary differential equations I: Nonstiff problems. Springer (1993)
 31. He, H., Bai, Y., Zhang, J., He, Q., Chen, H., Gan, Z., Wang, C., Li, X., Tian, G., Xie, L.: Mambaad: Exploring state space models for multi-class unsupervised anomaly detection. In: Globersons, A., Mackey, L., Belgrave, D., Fan, A., Paquet, U., Tomczak, J.M., Zhang, C. (eds.) Advances in Neural Information Processing Systems 38: Annual Conference on Neural Information Processing Systems 2024, NeurIPS 2024, Vancouver, BC, Canada, December 10 - 15, 2024 (2024), http://papers.nips.cc/paper_files/paper/2024/hash/833b21da1956c6b92f6df253bf655cf5-Abstract-Conference.html
 32. He, H., Zhang, J., Chen, H., Chen, X., Li, Z., Chen, X., Wang, Y., Wang, C., Xie, L.: A diffusion-based framework for multi-class anomaly detection. In: Wooldridge, M.J., Dy, J.G., Natarajan, S. (eds.) Thirty-Eighth AAAI Conference on Artificial

- Intelligence, AAAI 2024, Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence, IAAI 2024, Fourteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2014, February 20-27, 2024, Vancouver, Canada. pp. 8472–8480. AAAI Press (2024). <https://doi.org/10.1609/AAAI.V38I8.28690>, <https://doi.org/10.1609/aaai.v38i8.28690>
33. He, H., Zhang, J., Tian, G., Wang, C., Xie, L.: Learning multi-view anomaly detection with efficient adaptive selection. *IEEE Transactions on Multimedia* (2026)
 34. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016, Las Vegas, NV, USA, June 27-30, 2016. pp. 770–778. IEEE Computer Society (2016). <https://doi.org/10.1109/CVPR.2016.90>, <https://doi.org/10.1109/CVPR.2016.90>
 35. Heckler, L., König, R., Bergmann, P.: Exploring the importance of pretrained feature extractors for unsupervised anomaly detection and localization. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023 - Workshops, Vancouver, BC, Canada, June 17-24, 2023. pp. 2917–2926. IEEE (2023). <https://doi.org/10.1109/CVPRW59228.2023.00293>, <https://doi.org/10.1109/CVPRW59228.2023.00293>
 36. Heckler-Kram, L., Neudeck, J.H., Scheler, U., König, R., Steger, C.: The mvtec ad 2 dataset: Advanced scenarios for unsupervised anomaly detection. *arXiv preprint arXiv:2503.21622* (2025)
 37. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). *arXiv preprint arXiv:1606.08415* (2016)
 38. Hutchinson, M.F.: A stochastic estimator of the trace of the influence matrix for laplacian smoothing splines. *Communications in Statistics-Simulation and Computation* **18**(3), 1059–1076 (1989)
 39. Jeong, J., Zou, Y., Kim, T., Zhang, D., Ravichandran, A., Dabeer, O.: Winclip: Zero-/few-shot anomaly classification and segmentation. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023. pp. 19606–19616. IEEE (2023). <https://doi.org/10.1109/CVPR52729.2023.01878>, <https://doi.org/10.1109/CVPR52729.2023.01878>
 40. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), <http://arxiv.org/abs/1412.6980>
 41. Kingma, D.P., Dhariwal, P.: Glow: Generative flow with invertible 1x1 convolutions. In: Bengio, S., Wallach, H.M., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. pp. 10236–10245 (2018), <https://proceedings.neurips.cc/paper/2018/hash/d139db6a236200b21cc7f752979132d0-Abstract.html>
 42. Kirichenko, P., Izmailov, P., Wilson, A.G.: Why normalizing flows fail to detect out-of-distribution data. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (eds.) *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual* (2020), <https://proceedings.neurips.cc/paper/2020/hash/ecb9fe2fbb99c31f567e9823e884dbec-Abstract.html>

43. Kruse, M., Rosenhahn, B.: Multi-flow: Multi-view-enriched normalizing flows for industrial anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2025, Nashville, TN, USA, June 11-15, 2025. pp. 3933–3944. Computer Vision Foundation / IEEE (2025)
44. Kruse, M., Rudolph, M., Woiwode, D., Rosenhahn, B.: Splatpose & detect: Pose-agnostic 3d anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024 - Workshops, Seattle, WA, USA, June 17-18, 2024. pp. 3950–3960. IEEE (2024). <https://doi.org/10.1109/CVPRW63382.2024.00399>, <https://doi.org/10.1109/CVPRW63382.2024.00399>
45. Lee, S., Lee, S., Song, B.C.: CFA: coupled-hypersphere-based feature adaptation for target-oriented anomaly localization. *IEEE Access* **10**, 78446–78454 (2022). <https://doi.org/10.1109/ACCESS.2022.3193699>, <https://doi.org/10.1109/ACCESS.2022.3193699>
46. Lei, J., Hu, X., Wang, Y., Liu, D.: Pyramidflow: High-resolution defect contrastive localization using pyramid normalizing flow. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14143–14152 (2023)
47. Lipman, Y., Chen, R.T.Q., Ben-Hamu, H., Nickel, M., Le, M.: Flow matching for generative modeling. In: The Eleventh International Conference on Learning Representations, ICLR 2023, Kigali, Rwanda, May 1-5, 2023. OpenReview.net (2023), <https://openreview.net/forum?id=PqvMRDCJT9t>
48. Lipman, Y., Havasi, M., Holderrieth, P., Shaul, N., Le, M., Karrer, B., Chen, R.T., Lopez-Paz, D., Ben-Hamu, H., Gat, I.: Flow matching guide and code. arXiv preprint arXiv:2412.06264 (2024)
49. Liu, Y., Xu, X., Li, S., Liao, J., Yang, X.: Multi-view industrial anomaly detection with epipolar constrained cross-view fusion. arXiv preprint arXiv:2503.11088 (2025)
50. Liu, Z., Zhou, Y., Xu, Y., Wang, Z.: Simplenet: A simple network for image anomaly detection and localization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 20402–20411 (2023)
51. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. In: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net (2019), <https://openreview.net/forum?id=Bkg6RiCqY7>
52. McKay, M.D., Beckman, R.J., Conover, W.J.: A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **42**(1), 55–61 (2000)
53. Morokoff, W.J., Caflisch, R.E.: Quasi-monte carlo integration. *Journal of computational physics* **122**(2), 218–230 (1995)
54. Papamakarios, G., Nalisnick, E., Rezende, D.J., Mohamed, S., Lakshminarayanan, B.: Normalizing flows for probabilistic modeling and inference. *The Journal of Machine Learning Research* **22**(1), 2617–2680 (2021)
55. Rezende, D., Mohamed, S.: Variational inference with normalizing flows. In: International conference on machine learning. pp. 1530–1538. PMLR (2015)
56. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14318–14328 (2022)
57. Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but differnet: Semi-supervised defect detection with normalizing flows. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 1907–1916 (2021)

58. Rudolph, M., Wehrbein, T., Rosenhahn, B., Wandt, B.: Fully convolutional cross-scale-flows for image-based defect detection. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1088–1097 (2022)
59. Rudolph, M., Wehrbein, T., Rosenhahn, B., Wandt, B.: Asymmetric student-teacher networks for industrial anomaly detection. In: WACV. pp. 2591–2601. IEEE (2023)
60. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khali-dov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al.: Dinov3. arXiv preprint arXiv:2508.10104 (2025)
61. Tien, T.D., Nguyen, A.T., Tran, N.H., Huy, T.D., Duong, S., Nguyen, C.D.T., Truong, S.Q.: Revisiting reverse distillation for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 24511–24520 (2023)
62. Tong, A., Fatras, K., Malkin, N., Huguet, G., Zhang, Y., Rector-Brooks, J., Wolf, G., Bengio, Y.: Improving and generalizing flow-based generative models with minibatch optimal transport. *Trans. Mach. Learn. Res.* **2024** (2024), <https://openreview.net/forum?id=CD9Snc73AW>
63. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. *Advances in neural information processing systems* **30** (2017)
64. Wang, C., Zhu, W., Gao, B., Gan, Z., Zhang, J., Gu, Z., Qian, S., Chen, M., Ma, L.: Real-iad: A real-world multi-view dataset for benchmarking versatile industrial anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16–22, 2024. pp. 22883–22892. IEEE (2024). <https://doi.org/10.1109/CVPR52733.2024.02159>, <https://doi.org/10.1109/CVPR52733.2024.02159>
65. Wang, G., Han, S., Ding, E., Huang, D.: Student-teacher feature pyramid matching for anomaly detection. In: British Machine Vision Conference (2021)
66. Yang, K., Cao, J., Bai, Z., Su, Z., Tagliasacchi, A.: Piad: Pose and illumination agnostic anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 4734–4743 (June 2025)
67. You, Z., Cui, L., Shen, Y., Yang, K., Lu, X., Zheng, Y., Le, X.: A unified model for multi-class anomaly detection. *Advances in Neural Information Processing Systems* **35**, 4571–4584 (2022)
68. Yu, J., Zheng, Y., Wang, X., Li, W., Wu, Y., Zhao, R., Wu, L.: Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows. arXiv preprint arXiv:2111.07677 (2021)
69. Zagoruyko, S., Komodakis, N.: Wide residual networks. In: Wilson, R.C., Hancock, E.R., Smith, W.A.P. (eds.) Proceedings of the British Machine Vision Conference 2016, BMVC 2016, York, UK, September 19–22, 2016. BMVA Press (2016), <https://bmva-archive.org.uk/bmvc/2016/papers/paper087/index.html>
70. Zavrtnik, V., Kristan, M., Skočaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8330–8339 (2021)
71. Zavrtnik, V., Kristan, M., Skočaj, D.: Dsr—a dual subspace re-projection network for surface anomaly detection. In: European conference on computer vision. pp. 539–554. Springer (2022)
72. Zeiler, M.D., Krishnan, D., Taylor, G.W., Fergus, R.: Deconvolutional networks. In: The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13–18 June 2010. pp. 2528–2535.

- IEEE Computer Society (2010). <https://doi.org/10.1109/CVPR.2010.5539957>, <https://doi.org/10.1109/CVPR.2010.5539957>
73. Zhang, J., Chen, X., Wang, Y., Wang, C., Liu, Y., Li, X., Yang, M.H., Tao, D.: Exploring plain vit reconstruction for multi-class unsupervised anomaly detection. arXiv preprint arXiv:2312.07495 (2023)
 74. Zhang, J., He, H., Gan, Z., He, Q., Cai, Y., Xue, Z., Wang, Y., Wang, C., Xie, L., Liu, Y.: A comprehensive library for benchmarking multi-class visual anomaly detection. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 3524–3533 (2025)
 75. Zhang, J., Wang, G., Jin, Y., Huang, D.: Towards training-free anomaly detection with vision and language foundation models. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11–15, 2025. pp. 15204–15213. Computer Vision Foundation / IEEE (2025). <https://doi.org/10.1109/CVPR52734.2025.01416>, https://openaccess.thecvf.com/content/CVPR2025/html/Zhang_Towards_Training-free_Anomaly_Detection_with_Vision_and_Language_Foundation_Models_CVPR_2025_paper.html
 76. Zhang, X., Xu, M., Zhou, X.: Realnet: A feature selection network with realistic synthetic anomaly for anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16–22, 2024. pp. 16699–16708. IEEE (2024). <https://doi.org/10.1109/CVPR52733.2024.01580>, <https://doi.org/10.1109/CVPR52733.2024.01580>
 77. Zhang, X., Li, S., Li, X., Huang, P., Shan, J., Chen, T.: Destseg: Segmentation guided denoising student-teacher for anomaly detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3914–3923 (2023)
 78. Zhou, C., Paffenroth, R.C.: Anomaly detection with robust deep autoencoders. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. pp. 665–674 (2017)
 79. Zhou, K., Cao, Y., Kim, T., Zhao, H., Dong, H., Ting, K.M., Zhu, Y.: Rad: A dataset and benchmark for real-life anomaly detection with robotic observations. arXiv preprint arXiv:2410.00713 (2024)
 80. Zhou, Q., Li, W., Jiang, L., Wang, G., Zhou, G., Zhang, S., Zhao, H.: PAD: A dataset and benchmark for pose-agnostic anomaly detection. In: Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S. (eds.) Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023 (2023), http://papers.nips.cc/paper_files/paper/2023/hash/8bc5aef775aacc1650a9790f1428bcea-Abstract-Datasets_and_Benchmarks.html
 81. Zhou, Q., Pang, G., Tian, Y., He, S., Chen, J.: Anomalyclip: Object-agnostic prompt learning for zero-shot anomaly detection. In: The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7–11, 2024. OpenReview.net (2024), <https://openreview.net/forum?id=buC4E91xZE>
 82. Zhu, W., Wang, L., Zhou, Z., Wang, C., Pan, Y., Zhang, R., Chen, Z., Cheng, L., Gao, B., Zhang, J., Gan, Z., Wang, Y., Chen, Y., Qian, S., Chi, M., Peng, B., Ma, L.: Real-iad D3: A real-world 2d/pseudo-3d/3d dataset for industrial anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11–15, 2025. pp. 15214–15223. Computer Vision Foundation / IEEE (2025). <https://doi.org/10.1109/CVPR52734.2025.01417>, https://openaccess.thecvf.com/content/CVPR2025/html/Zhu_Real-

IAD_D3_A_Real-World_2DPseudo-3D3D_Dataset_for_Industrial_Anomaly_Detection_CVPR_2025_paper.html

83. Zou, Y., Jeong, J., Pemula, L., Zhang, D., Dabeer, O.: Spot-the-difference self-supervised pre-training for anomaly detection and segmentation. In: European Conference on Computer Vision. pp. 392–408. Springer (2022)