IPD: An Industrial Pipeline Dataset for Anomaly Detection and Localization

1st Diamantis Rafail Papadam
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
drpapadam@csd.auth.gr

2nd Pantelis Mentesidis
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
mentpant@csd.auth.gr

3rd Alexandros Zamioudis
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
azamiou@csd.auth.gr

4th Vasileios Mygdalis
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
mygdalisv@csd.auth.gr

5th Dimitrios Psarras
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
dpsarras@csd.auth.gr

6th Ioannis Pitas
Department of Informatics
Aristotle University of Thessaloniki
Thessaloniki, Greece
pitas@csd.auth.gr

***Abstract*—This paper examines the challenging task of visual inspection on industrial pipelines. Key challenges in this area include: i) highly noisy and cluttered pipeline background, ii) damage types are not known a-priori, iii) high variability in damage visual appearance, and iv) imbalanced distribution of the training samples (images of damaged pipelines appear less frequently than non-damaged). Therefore, visual Anomaly Detection (AD) and Anomaly Localization (AL) are valid and suitable methods to address such a task. This paper introduces the Industrial Pipeline Dataset (IPD), that is mainly collected from industrial facilities in Europe and focuses on damages to insulated pipelines. Data collection and annotations was guided by consultations with industry experts to ensure the accurate identification of insulation damages. State-of-the-art AD and AL methods were evaluated to establish a baseline for the inspection task. Their rather mediocre performance, shown both visually and empirically, showcases the complexity of the task. As such, this paper makes two main contributions: i) IPD is, to the best of our knowledge, the first publicly available AD/AL dataset on visual inspection of industrial pipeline insulation, making it highly valuable for real-world industrial applications, and ii) since modern AD/AL methods fail to perform well on IPD, it constitutes a valuable research benchmark.**

***Index Terms*—Anomaly Detection, Anomaly Localization, Background Noise, Industrial Pipelines, Dataset**

I. Introduction

Inspection of industrial pipelines in large facilities is usually performed by examining the internal condition of the pipelines using specialized tools (*e.g., borescopes*) [1]. These methods perform remarkably well, but they have two serious disadvantages: i) requiring (parts of) the factory to be shut down, which in the oil and gas industry may translate to reduction of production capacity, resulting in financial repercussion during the inspection day, and ii) damages to the insulation cannot be detected. In our research efforts, we consider methodologies to inspect the pipeline and its insulation during normal operation. As such, visual inspection from the outside of the pipeline is usually the first option to be considered. In order to avoid costs related to scaffolding and dangers related to worker exposure to high altitudes, drones are commonly used to collect videos of the pipeline that are sent to an inspector. Hence, video stream analysis can be automated by machine learning methods that detect damage in pipeline insulation.

This work examines the visual insulation inspection automation problem from the perspectives of visual Anomaly Detection (AD) & Localization (AL). We refer to AD as the binary classification task at the image level, while we consider AL to be the pixel-level classification task (or segmentation) of the anomaly within the image [2]. In practice, a data sample is considered anomalous if it does not conform to some definition of normality [3]. For example, in image reconstruction approaches [4], normality is defined by selecting an acceptable threshold for the L_2 distance between the initial and reconstructed images. In other approaches [5], normality is defined by the probability of an image patch to belong to the multivariate Gaussian probability distribution that is above a chosen threshold.

AD and AL are actively being used in many important applications (*e.g., defect detection, quality control, medical imaging*) [6]. However, in the industrial domain, there is a serious lack of publicly available AD/AL ground-truth data [7]. Even more so, when pixel-level annotations are needed. Two well-established datasets for industrial AD/AL are MVTecAD [8] and BTAD [6], where the images are captured in nearly ideal conditions, with well-centered objects and practically no background noise. VisA [9] is a slightly more complex dataset that is also used as a benchmark. However, all these datasets are still very simple, as several AD/AL approaches achieve a near-perfect performance on them [10], [11].

This paper introduces the Industrial Pipeline Dataset (IPD[1]). Unlike existing AD/AL datasets, IPD contains images of insulated pipelines—mainly from industrial facilities located in Europe—with high background noise and clutter, as well

[1]The IPD dataset is available at: https://aiia.csd.auth.gr/ipd-dataset/

as other challenging aspects that are presented in Section III. We have employed state-of-the-art (SOTA) AD/AL methods, showcasing their quite poor performance on IPD. Therefore, the proposed dataset is a valuable research benchmark to help advance the AD/AL SOTA even further.

The rest of the paper is structured as follows. Section II presents well-established AD/AL datasets and algorithms that have been evaluated in the proposed IPD dataset. Section III thoroughly presents the IPD dataset and its feature distribution compared to existing AD/AL benchmarks. Section IV presents a detailed performance evaluation of several modern AD/AL approaches on IPD and showcases the difficulty of the proposed dataset. Finally, Section V summarizes the findings of this paper and presents future work directions.

II. Anomaly detection and localization

This section carries out a brief AD/AL literature review. In the first subsection, three popular industrial AD/AL datasets, typically used by unsupervised or semi-supervised approaches, are presented. The second subsection, presents the AD/AL algorithms—most of which are SOTA—that are evaluated on the IPD dataset in Section IV.

A. Anomaly Detection and Localization Datasets

MVTecAD [8] is a very popular AD/AL dataset, consisting of 15 classes with an average number of 357 images per class. It mainly focuses on industrial inspection, with 10 object classes (*e.g., cable, screw, transistor*), and 5 texture classes (*e.g., tile, wood*). However, it is very simplistic, since all objects are well centered and there is no background noise.

BTAD [6] is a similarly simple AD/AL benchmark that comprises 3 classes with an average of 847 images per class. Each class is an industrial product, though it's not specified exactly what it is.

VisA [9] is a bit more complicated compared to the above-mentioned datasets in two main ways: i) multiple object instances appear in a single image, and ii) objects are not always centered and may be located in various image parts. The VisA dataset consists of 12 classes with an average of 902 images per class.

B. Anomaly Detection and Localization Methods

The methods that this paper evaluates on IPD are either unsupervised or semi-supervised, meaning that they do not learn from anomalous samples during training, unless they synthetically construct such samples (semi-supervised learning, *e.g., Draem* [12]). According to a recent survey [7], current industrial AD/AL research is split in two main categories: i) feature-embedding based methods, and ii) reconstruction based methods. More specifically, in our paper we are testing methods that belong in the following sub-categories: i_α) Teacher-Student (T-S) architecture, i_β) distribution map, i_γ) memory bank, and ii_α) autoencoder.

T-S methods: Following this taxonomy, let us begin with the T-S architecture-based methods, and particularly *STFPM* [13]. The teacher DNN is a pre-trained residual network (ResNet), and the Student DNN has the same architecture as the Teacher. During training, non-anomalous images are given as input to the Student DNN, while it learns to match the Teacher feature maps at different levels. During testing, the multi-level feature map difference of the two networks is used to determine the anomaly score. In a similar fashion, *Reverse Distillation* [14] adopts the T-S paradigm but with an interesting novelty. Instead of the raw image, the Student network receives as input the Teacher's multiscale feature representation, projected in a latent space, and learns to reconstruct it using normal samples. The so called one-class bottleneck embedding (OCBE) module, which projects the Teacher's feature maps into the latent space, is also trained along with the student network.

Distribution map methods: To move to approaches that are based on distribution maps, in *Rkde* [15], Faster R-CNN is used for region proposals [16], followed by *AlexNet* [17] feature extraction. As a final step, Kernel Density Estimation (KDE) is performed to model the normal features and detect anomalies during inference. In *DFM* [18], the probability distribution of the features extracted by a pre-trained classification DNN is modeled by: a Gaussian distribution, or a Gaussian mixture model. This probability density function (pdf) model is used to perform Out-Of-Distribution (OOD) detection at inference time. More modern AD approaches incorporate Normalizing Flows (NFs) [19]. In *Cflow* [20], image patches are fed into a CNN feature extractor. The patch feature maps are used for normal-sample distribution modeling. These feature maps, along with the patch positional encodings are passed as input in NF decoders that transform the feature probability distribution into a Gaussian one. In a slightly different manner, *FastFlow* [21] models the feature probability distribution by a two-dimensional normal pdf. *Uflow* [22] employs a *U-Net* architecture [23], where the multi-scale feature extractor acts like the encoder, and a set of Normalizing Flows acts as the decoder. Furthermore, each NF is trained to map the normal image data to a Gaussian pdf. At inference time, anomaly maps are created at each feature scale, and they get upsampled to the original image size. Moreover, an estimate of the Number of False Alarms (*NFA*) [24], is used to produce the final AL mask. In *CSflow* [25], the input is processed at 3 scales in parallel. Moreover, the vanilla NF is extended with a novel cross-scale flow that allows feature maps of different sizes to interact with each other during training and inference.

Memory bank methods: Moving to approaches that are based on memory banks, in *Padim* [5], the input is indirectly split into image patches, by considering the corresponding areas of a pre-trained ResNet's feature map. A Gaussian distribution is created for each patch separately, by considering feature maps at different ResNet levels. This patch-level pdf creation with normal samples, which is stored in a memory bank, allows for AL when an abnormal image is fed into the network. In *Patchcore* [26], image patches are considered in a similar fashion to *Padim*, by passing the images into a pre-trained ResNet for feature extraction. The extracted features are subsampled and stored in a memory bank. During inference, the anomaly score for a test patch is calculated by performing

a nearest neighbor search between the test sample features and the features stored in the memory bank. In *CFA* [27], the authors incorporate a patch descriptor and a memory bank, to adapt the features to the target dataset. As such, during training, they define a coupled-hypersphere, and during inference, they examine whether a test sample is within the constructed hypersphere (normal), or it is outside (abnormal).

Autoencoder methods: To conclude this section, we present reconstruction-based approaches that use autoencoders. In *Dsr* [28], an architecture based on a quantized feature space representation is proposed. Moreover, a specific object appearance and a general object appearance decoder are used in parallel to reconstruct the input image. Finally, an anomaly detection module combines the two reconstructions and an upsampling module is used to produce the final AL mask. In *Draem* [12], a semi-supervised approach is followed, using synthetic anomaly creation. The proposed architecture incorporates a reconstructive and a discriminative network used in sequence. During training, each network is trained with a different loss function. Finally, during testing, the discriminative network that takes as input the concatenation of the initial and reconstructed images, and produces the AL predicted mask.

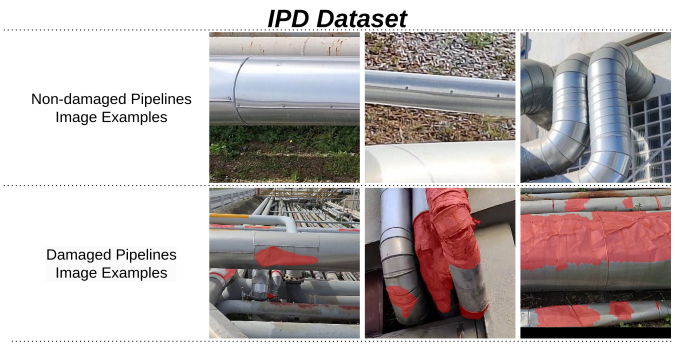


Fig. 1: IPD image examples. First row depicts non-damaged pipeline images (*i.e.*, normal samples). Second row depicts damages pipeline images (*i.e.*, abnormal samples), with pixel-level damage annotations (shown in red color).

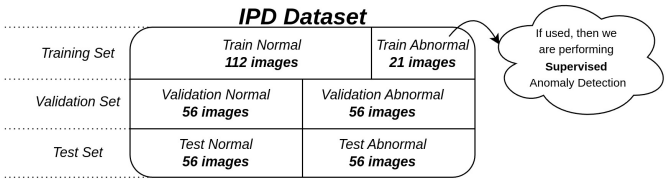


Fig. 2: IPD split for AD and AL tasks.

III. IPD Dataset

In this section, the IPD dataset is presented in detail. Moreover, we showcase descriptive statistics of IPD, and also compare its DINOv2 extracted features with three other well-established AD/AL benchmarks.

Data collection: The dataset was mainly collected from industrial facilities in Europe, focusing on insulated that are jacketed with metal. To increase the size of the dataset, images with such pipelines were also captured from a public facility. Data collection was guided by consultations with industry experts to ensure the accurate identification of insulation damages. Initially, the focus was on more visible defects, such as i) open insulations and ii) holes. However, expert recommendations highlighted the importance of also considering iii) dents in the insulation as potential indicators of damage, as they may suggest a defect in others parts of the pipeline. All damaged areas were labeled with a pixel value of 1, meaning that there is no distinction between damage types in the annotations. A professional camera and two modern smartphones were used to capture the images in 5 different resolutions: i) 2353×4080 (9.6 MP), ii) 3000×4000 (12 MP), iii) 3060×4080 (12.5 MP), iv) 4000×6000 (24 MP), v) 6120×8160 (49.9 MP).

Data pre-processing: To maintain the confidentiality of the facilities, areas that could potentially reveal the identity of specific plants removed from the dataset. Furthermore, to increase the size of the dataset as well as the variance of image resolutions, we performed some manual croppings, effectively replacing large images with several smaller ones.

Quality assurance: To ensure that the quality of the annotations is good, we did not only consult experts as mentioned previously, but also made sure that 5 colleagues collaborate in the annotation process. These colleagues worked together, peer reviewed each other, and went over a few iterations of updating the annotations to achieve the final IPD dataset AD and AL ground-truth labels and masks respectively. Figure 1, illustrates samples from the annotated IPD dataset, where the AL (pixel-level) ground-truth masks are presented for damaged pipelines.

Dataset splits: The IPD dataset consists of 357 images that contain insulated pipelines. As depicted in Figure 2, the training set consists of 112 normal images and an optional subset of 21 abnormal images, which can be used by supervised learning approaches, but it must be omitted by unsupervised and semi-supervised ones. Moreover, the validation and test sets both consist of 56 normal and 56 abnormal images.

Dataset statistics and comparison with relevant AD/AL benchmarks: Compared to current well-established AD/AL benchmarks (*e.g., MVTecAD, BTAD, VisA), IPD* has the following challenging aspects: i) the pipeline (object) background is highly noisy and cluttered, ii) some pipelines (objects) take up a big portion of image pixels while others a tiny portion, and iii) there is a high variance of image resolutions in the dataset. Figure 3, illustrates the resolution distribution of images in IPD. Finally, Figure 4, shows the DINOv2 feature distribution—after performing t-SNE [29]—of the IPD dataset with respect to MVTecAD, BTAD, and VisA datasets (where each point corresponds to a single image), highlighting that in the IPD dataset image features exhibit greater variance. This is attributed to the complexity of the dataset.

IV. Experiments on IPD Dataset

Implementation and metrics: Several unsupervised/semi-supervised AD/AL algorithms—most of which constitute the SOTA—were evaluated on IPD, using the anomalib framework v1.1.1 [30]. All images were resized to 256x256 resolution.

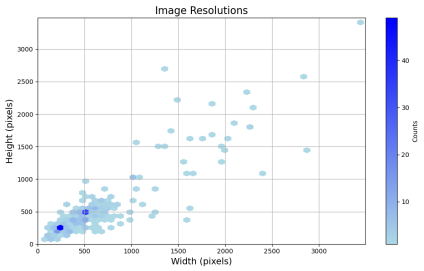
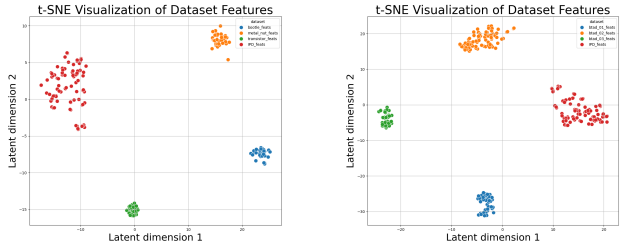
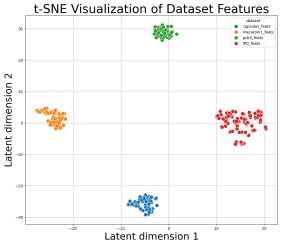


Fig. 3: IPD image resolution distribution.



(a) MVTecAD – IPD

(b) BTAD – IPD



(c) VisA – IPD

Fig. 4: AD/AL dataset DINOv2 feature visualization. (a) Three classes of MVTecAD with respect to IPD. (b) All three classes of BTAD with respect to IPD. (c) Three classes of VisA with respect to IPD.

The *Recall*, *Precision*, *Accuracy*, *F1 Score*, and *AUROC* metrics were used for AD/AL evaluation. To conduct the experiments in a more reliable manner, 5 different seeds were used and the results were averaged to get the final performance metric evaluations. The data split into training, validation, and test sets remained unchanged throughout all experiments, and it is the same as in the IPD dataset we are publishing. No fine-tuning was performed, and the experiments were conducted with the default method hyperparameters as defined in the anomalib framework v1.1.1 [30], except for the number of epochs, which was set to 100 where applicable.

Results and key findings: In Table I, the AD (image-level) results are reported, while in Table II, the AL (pixel-level) results are reported. As *AUROC* is considered to be the most informative performance metric, the rows in both tables are sorted in decreasing *AUROC* order. We use **bold style** to denote the best performance with respect to each metric, and underline style to denote the second-best performance. As a main observation, we report the mediocre performance of even the best AD/AL algorithms on the IPD dataset. This proves our point, that more challenging datasets, like IPD, should be published and used as benchmarks to enhance research in the AD/AL domain. Moreover, it is observed that for the AD task, the best 2 approaches (*i.e.*, *Dfm*, *Uflow*) are based on distributions maps. Regarding the AL task, we notice that the best approach (*i.e.*, *Patchcore*) is based on a memory bank, and the second-best approach (*i.e.*, *Uflow*) is based on a distribution map. Moreover, *Reverse Distillation*, which is based on a T-S architecture, achieves 3^{rd} place in both tasks. These observations reveal the fact there is currently not a single type of AD/AL approach that clearly outperforms the others. Furthermore, in Figure 5, a few inference results of the best-performing AL methods are visualized, highlighting the inability of current SOTA methods to successfully detect and localize anomalies on IPD, while they perform remarkably well on the MVTecAD dataset.

TABLE I: AD Performance on IPD: image-level evaluation

Algorithm	Family	*AUROC*	*Accuracy*	*F1 Score*	*Precision*	*Recall*
Dfm	Distribution map	**0.816**	**0.821**	**0.836**	**0.773**	0.911
Uflow	Distribution map	0.811	0.759	0.797	0.690	0.943
Reverse Distillation	Teacher-Student architecture	0.782	0.711	0.752	0.658	0.879
Patchcore	Memory bank	0.777	0.725	0.755	0.681	0.846
STFPM	Teacher-Student architecture	0.759	0.684	0.737	0.645	0.879
Fastflow	Distribution map	0.736	0.636	0.719	0.587	0.929
Draem	Autoencoder	0.718	0.582	0.691	0.559	0.929
Dsr	Autoencoder	0.668	0.641	0.723	0.599	0.925
Padim	Memory bank	0.652	0.602	0.695	0.564	0.907
Rkde	Distribution map	0.639	0.518	0.675	0.509	**1.000**
Cfa	Memory bank	0.594	0.514	0.669	0.508	0.982
Csflow	Distribution map	0.544	0.566	0.698	0.556	0.968
Cflow	Distribution map	0.341	0.407	0.548	0.441	0.732

TABLE II: AL Performance on IPD: pixel-level evaluation

Algorithm	Family	*AUROC*	*Accuracy*	*F1 Score*	*Precision*	*Recall*
Patchcore	Memory bank	**0.798**	**0.743**	**0.470**	**0.360**	0.677
Uflow	Distribution map	0.794	0.728	0.467	0.349	0.709
Reverse Distillation	Teacher-Student architecture	0.780	0.715	0.454	0.335	0.705
STFPM	Teacher-Student architecture	0.714	0.682	0.388	0.287	0.598
Padim	Memory bank	0.711	0.652	0.391	0.277	0.664
Dfm	Distribution map	0.706	0.611	0.401	0.270	0.777
Fastflow	Distribution map	0.623	0.551	0.332	0.222	0.660
Rkde	Distribution map	0.601	0.537	0.343	0.225	0.719
Cflow	Distribution map	0.581	0.578	0.280	0.198	0.498
Draem	Autoencoder	0.577	0.376	0.314	0.196	0.833
Csflow	Distribution map	0.550	0.478	0.302	0.208	0.649
Dsr	Autoencoder	0.469	0.168	0.287	0.168	**0.999**
Cfa	Memory bank	0.465	0.284	0.294	0.177	0.884

V. Conclusions and Future Work

In this paper, we presented the new IPD dataset which is, to the best of our knowledge, the first public AD/AL dataset comprising RGB images of insulated industrial pipelines. Moreover, we evaluated modern AD/AL algorithms—most of which are SOTA—on IPD, and verified both visually and by using well-established metrics, such as *AUROC*, that there is ample room for AD/AL algorithm improvement. As such, IPD constitutes a challenging and useful benchmark for the AD/AL tasks. Moreover, it is very useful for the improvement of automation in industrial applications. Regarding future work, we intend to: i) perform a baseline evaluation by using IPD as

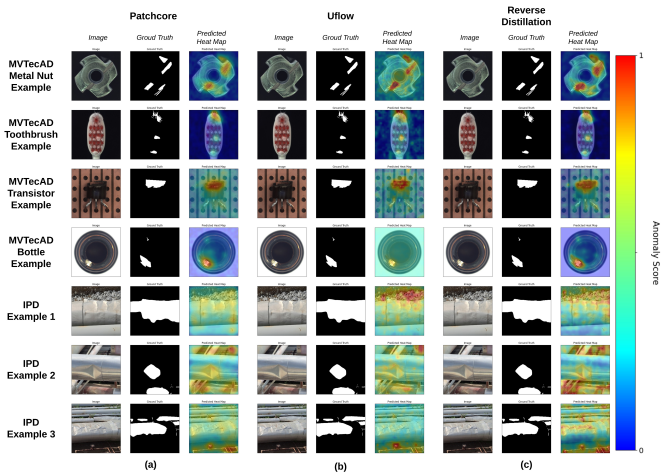


Fig. 5: IPD inference examples using: (a) *Patchcore*, (b) *Uflow*, and (c) *Reverse Distillation*. The methods can detect and localize damages very well on MVTecAD samples, while they clearly struggle on the IPD dataset.

an object detection dataset, where the objects are the damages, $ii)$ annotate the pipelines of IPD and perform a pipeline semantic segmentation baseline evaluation, and $iii)$ increase the size of the IPD dataset.

ACKNOWLEDGEMENTS

This work has received funding from the European Union's Horizon Europe research and innovation programme under grant agreement number 101070604 (SIMAR). This publication reflects only the authors' views. The European Commission is not responsible for any use that may be made of the information it contains.

REFERENCES

[1] D. Yang, C. Ma, G. Yu, and Y. Chen, "Automatic defect detection of pipelines based on improved ofg-yolo algorithm," *Measurement*, vol. 242, p. 115847, 2025.

[2] X. Tao, X. Gong, X. Zhang, S. Yan, and C. Adak, "Deep learning for unsupervised anomaly localization in industrial images: A survey," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–21, 2022.

[3] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[4] V. Zavrtanik, M. Kristan, and D. Skočaj, "Reconstruction by inpainting for visual anomaly detection," *Pattern Recognition*, vol. 112, p. 107706, 2021.

[5] T. Defard, A. Setkov, A. Loesch, and R. Audigier, "Padim: a patch distribution modeling framework for anomaly detection and localization," in *International Conference on Pattern Recognition*. Springer, 2021, pp. 475–489.

[6] P. Mishra, R. Verk, D. Fornasier, C. Piciarelli, and G. L. Foresti, "Vt-adl: A vision transformer network for image anomaly detection and localization," in *2021 IEEE 30th International Symposium on Industrial Electronics (ISIE)*. IEEE, 2021, pp. 01–06.

[7] J. Liu, G. Xie, J. Wang, S. Li, C. Wang, F. Zheng, and Y. Jin, "Deep industrial image anomaly detection: A survey," *Machine Intelligence Research*, vol. 21, no. 1, pp. 104–135, 2024.

[8] P. Bergmann, M. Fauser, D. Sattlegger, and C. 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*, 2019, pp. 9592–9600.

[9] Y. Zou, J. Jeong, L. Pemula, D. Zhang, and O. Dabeer, "Spot-the-difference self-supervised pre-training for anomaly detection and segmentation," in *European Conference on Computer Vision*. Springer, 2022, pp. 392–408.

[10] S. Damm, M. Laszkiewicz, J. Lederer, and A. Fischer, "Anomalydino: Boosting patch-based few-shot anomaly detection with dinov2," *arXiv preprint arXiv:2405.14529*, 2024.

[11] H. Zhang, Z. Wang, Z. Wu, and Y.-G. Jiang, "Diffusionad: Norm-guided one-step denoising diffusion for anomaly detection," *arXiv preprint arXiv:2303.08730*, 2023.

[12] V. Zavrtanik, M. Kristan, and D. Skočaj, "Draem-a discriminatively trained reconstruction embedding for surface anomaly detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 8330–8339.

[13] G. Wang, S. Han, E. Ding, and D. Huang, "Student-teacher feature pyramid matching for anomaly detection," *arXiv preprint arXiv:2103.04257*, 2021.

[14] H. Deng and X. Li, "Anomaly detection via reverse distillation from one-class embedding," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9737–9746.

[15] P. Adey, O. Hamilton, M. Bordewich, and T. Breckon, "Region based anomaly detection with real-time training and analysis," in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, pp. 495–499.

[16] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *Advances in neural information processing systems*, vol. 28, 2015.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Advances in neural information processing systems*, vol. 25, 2012.

[18] N. A. Ahuja, I. Ndiour, T. Kalyanpur, and O. Tickoo, "Probabilistic modeling of deep features for out-of-distribution and adversarial detection," *arXiv preprint arXiv:1909.11786*, 2019.

[19] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, "Normalizing flows for probabilistic modeling and inference," *Journal of Machine Learning Research*, vol. 22, no. 57, pp. 1–64, 2021.

[20] D. Gudovskiy, S. Ishizaka, and K. Kozuka, "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*, 2022, pp. 98–107.

[21] J. Yu, Y. Zheng, X. Wang, W. Li, Y. Wu, R. Zhao, and L. Wu, "Fastflow: Unsupervised anomaly detection and localization via 2d normalizing flows," *arXiv preprint arXiv:2111.07677*, 2021.

[22] M. Tailanian, Á. Pardo, and P. Musé, "U-flow: A u-shaped normalizing flow for anomaly detection with unsupervised threshold," *Journal of Mathematical Imaging and Vision*, pp. 1–19, 2024.

[23] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[24] A. Desolneux, L. Moisan, and J.-M. Morel, *From gestalt theory to image analysis: a probabilistic approach*. Springer Science & Business Media, 2007, vol. 34.

[25] M. Rudolph, T. Wehrbein, B. Rosenhahn, and B. Wandt, "Fully convolutional cross-scale-flows for image-based defect detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2022, pp. 1088–1097.

[26] K. Roth, L. Pemula, J. Zepeda, B. Schölkopf, T. Brox, and P. Gehler, "Towards total recall in industrial anomaly detection," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14 318–14 328.

[27] S. Lee, S. Lee, and B. C. Song, "Cfa: Coupled-hypersphere-based feature adaptation for target-oriented anomaly localization," *IEEE Access*, vol. 10, pp. 78 446–78 454, 2022.

[28] V. Zavrtanik, M. Kristan, and D. Skočaj, "Dsr – a dual subspace re-projection network for surface anomaly detection," in *Computer Vision – ECCV 2022*, S. Avidan, G. Brostow, M. Cissé, G. M. Farinella, and T. Hassner, Eds. Cham: Springer Nature Switzerland, 2022, pp. 539–554.

[29] L. Van der Maaten and G. Hinton, "Visualizing data using t-sne." *Journal of machine learning research*, vol. 9, no. 11, 2008.

[30] S. Akcay, D. Ameln, A. Vaidya, B. Lakshmanan, N. Ahuja, and U. Genc, "Anomalib: A deep learning library for anomaly detection," in *2022 IEEE International Conference on Image Processing (ICIP)*, 2022, pp. 1706–1710.