Self-supervised Learning for Robust Surface Defect Detection

Muhammad Aqeel(✉), Shakiba Sharifi, Marco Cristani, and Francesco Setti

Department of Engineering for Innovation Medicine, University of Verona,
Strada le Grazie 15, Verona, Italy
muhammad.aqeel@univr.it

Abstract. In this study, we discuss about the use of Self-Supervised Learning to improve robustness of Surface Defect Detection (SDD) models. We show how different state-of-the-art SDD methods are already implementing some sort of self-supervision in their learning procedure, and we discuss how more advanced techniques inspired to Confident Learning can be used in a generic pipeline. We also propose One-Shot Removal strategy, a baseline approach that can be applied to any SDD model to improve its robustness. Our method employs a three-step training pipeline: initial training on the entire dataset, followed by removal of anomalous samples, and fine-tuning on the refined dataset. Experiments conducted on the challenging Kolektor SDD2 dataset show how this process enhances the representation of 'normal' data and mitigates overfitting risks.

Keywords: Self-Supervised learning · Robust anomaly detection · Surface defect detection · Confident learning

1 Introduction

Surface defect detection (SDD) is a crucial task in modern manufacturing industry, involving quality control on materials like marble [1], steel [2], and leather [3]. However, detecting surface anomalies is challenging due to the diversity and complexity of surface textures, the rarity of anomalies, and the scarcity of labeled data for supervised learning. Anomalies are usually localized, occupying a limited portion of the image, making it extremely difficult for both humans and machines to isolate them. Indeed, SDD is very challenging for human inspectors, leading to difficulties in accurately identifying surface anomalies. As a result, the process of labeling training data becomes laborious and susceptible to errors.

In the last few years, most of the research has focused on unsupervised anomaly detection. These methods rely on the assumption of clean datasets from which the classifier learns to distinguish between normal and anomalous samples. In this scenario, the model is fed with normal samples at training time, while anomalous samples can only be seen at testing time. This setup alleviates the need for labor-intensive annotation and reduces the problems related to bad noisy labels. Nevertheless, the construction of the training set remains a critical point. Indeed, to avoid the risk of introducing anomalies in the training data, the user can be tempted to select only extremely clean

A. Fred et al. (Eds.): DeLTA 2024, CCIS 2172, pp. 164–177, 2024.
https://doi.org/10.1007/978-3-031-66705-3_11

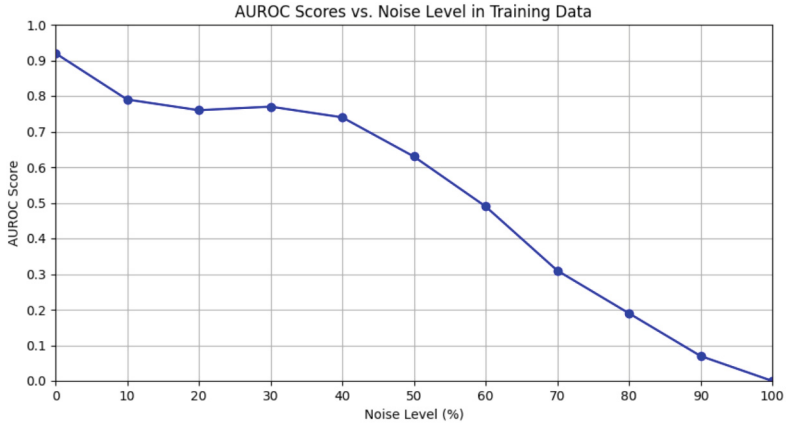

Fig. 1. Typical impact of noisy data in training an anomaly detection model.

samples; as a result, the trained model will likely overfit the data, generating at testing time a high number of false positives, *i.e.* normal samples classified as anomalies. On the other hand, including anomalous samples in the training set would generate a drop in performances due to the high number of false negatives. The typical behaviour of unsupervised anomaly detection models is shown in Fig. 1.

Therefore, there is a call for SDD models that are robust to noise in the labeling process, where samples may be incorrectly associated with the good or defective class. Recent developments in the field of deep learning, including autoencoders [4], generative adversarial networks [5], and discrete feature space representation [6], present encouraging prospects for addressing these challenges through the use of knowledge gained from unlabeled or weakly labeled data. Despite their potential, these approaches continue to encounter challenges including the determination of the most optimal anomaly score function, ensuring the level of robustness against noise and outliers, and the ability to generalize to data that is unseen [7,8]. Additional research is required to improve the efficacy and efficiency of robust surface anomaly detection methods. This position paper discusses the recent work related to the field of robust anomaly detection and presents a preliminary study proposing a simple yet effective method for handling noisy training data by leveraging self-supervision and confident learning literature. We propose a three-step training pipeline where *(1)* the model is trained on all the available data samples, then *(2)* evaluation is performed on the training data to isolate and remove anomalous samples, and finally *(3)* the model is fine-tuned using solely the filtered samples. This model not only shows improved performance on new, unseen data but also maintains increased resilience to adversarial samples, *i.e.* inputs that are intentionally designed to confuse the model, thereby ensuring its robustness against deliberately manipulated input that might distort the analysis promoting the learning of simpler and more generalizable patterns from the data. The main contributions of this paper can be summarized as follows:

– We present an approach for anomaly detection in presence of unreliable training data that uses a composite score that combines the likelihood of anomalies and deviation

magnitude to identify and remove influential outliers preserving valuable variability within the dataset;
- The One Shot Removal (OSR) technique validates samples based on a threshold value by systematically discarding high-scoring outliers, the training dataset is refined. This process allows the model to acquire a more representative set of features representing 'normal' conditions, thereby mitigating the risk of overfitting.
- The integration of One Shot Removal (OSR) with threshold value adjustment significantly reduces the model's complexity, enhancing its interpretability and reliability in detecting surface anomalies.

To summarize, our contribution, which includes the One Shot Removal technique and threshold value adjustment process, is an important advancement of the training process that improves the model's performance and efficiency, providing the way for more advanced robust surface defect detection (SDD) mechanisms. This novel technique defines new benchmark in the domain by providing a comprehensive solution to the problems of anomaly detection in a diverse range of industrial applications.

2 Related Work

2.1 Robust Anomaly Detection (RAD)

The general task of robust anomaly detection has been addressed over the years from several points of view, mostly targeting problems such as fraud detection [9] and intrusions in communication networks [10]. Various approaches have been explored, spanning from robust statistical methods [11] to deep learning approaches such as autoencoders [8,12] or recurrent neural networks [13]. The common aspect of all these methods is the fact they address the problem of RAD by developing methods that are less sensitive to noise in the labeling process. In this paper, we propose a different standpoint, that is, the idea that we can use whatever high-performing anomaly detector, focusing instead on the ability to remove as many noisy labels as possible. A similar approach is proposed in [14], where the authors propose a two-layer online learning framework for robust anomaly detection in the presence of unreliable anomaly labels, where the first layer filters out the suspicious data, and the second layer detects the anomaly patterns from the remaining data. This is very close to our approach, with the main difference that two different models are trained in this approach: one model is only devoted to predicting the quality of training data, and the second is the anomaly detector that will run at inference time. We argue that training the first model can become a difficult task itself, introducing a computational complexity that is possibly not compensated. Indeed, this framework is mostly used in online setups, where the temporal dimension can stabilize results and support the convergence of the label quality predictor.

2.2 Surface Defect Detection (SDD)

The process of detecting and identifying defects is essential in manufacturing to ensure a manufacturing process functions correctly and is under control. Surface defect detection includes the identification of scratches, blemishes, blockages from foreign objects,

discoloration, holes, and other irregularities on the surface of products [15]. Recent survey papers overview the state of the art in SDD from a technological and methodological point of view [16,17]. Traditionally, this field is divided between supervised methods, where both anomalies and normal samples are available during training, and unsupervised methods, where only normal samples can be used for training. While supervised methods have been proven to achieve better results on public benchmarks, the cost of producing these annotated datasets and the advancements in generative artificial intelligence have pushed the research community's interest toward unsupervised SDD. Many recent SDD methods are based on image reconstruction [18–20], where an encoder-decoder network is trained to reconstruct anomaly-free images. Comparison of original and reconstructed images will then allow us to spot anomalies. The underlying assumption is that these networks will be unable to accurately reconstruct anomalous regions because they have never seen them during training. These methods can produce a high false positive rate in case of high variability of training data. Instead of comparing the images, authors of [4] and [21] propose to perform this comparison at a feature level, thus learning to encode and decode features instead of whole images. Finally, discriminative unsupervised anomaly detection methods utilize synthetically generated anomalies to train a discriminative anomaly detection network. These methods tend to overfit the normal data, due to a limited distribution of the generated synthetic anomalies. [22] uses a reconstruction network to alleviate this problem by restoring the normal appearance of anomalous images. A different approach relies on the introduction of additional information in the form of textual prompts describing the defects as in [23].

2.3 Self-supervised Learning

Researchers have been investigating the potential of self-supervised learning frameworks for surface defect detection (SDD), leveraging unlabeled crack image datasets. [24] utilized an end-to-end training approach with a two-stage neural network, integrating SSL techniques to enhance the model's ability to learn from unlabeled data. They employed a unique loss function to address the uncertainty associated with region-based annotations, comparing it with precise pixel-level annotations. Similarly, [25] optimized a DCNN architecture for SSL, achieving notably higher crack detection accuracy on other datasets. Their SSL approach facilitated better generalization without considering the computational complexity and training period for online prediction. MaxPooling with a CNN was employed by [26] for SDD, incorporating SSL to compare its performance with traditional approaches. [27] utilized a DCC-Center Net architecture for SDD detection, employing SSL alongside keypoint estimation to identify center points and regress defect properties. However, the segmentation method and network architecture limited the model's effectiveness in detecting obscure defects.

Furthermore, [28] utilized a domain-adaptation adaptive CNN for SDD, employing SSL principles with adaptive learning rates based on loss and weight. Although the proposed model showed improved results compared to conventional CNN models, it was evaluated using a small dataset, indicating the need for further research with larger datasets to validate the SSL approach. Given these challenges, the application of confident learning could be instrumental in enhancing the robustness and reliability of SSL models in SDD. Very related to SSL, *confident learning* deals with the estimation of

uncertainty in dataset labels to understand the reliability or certainty of ground-truth labels associated with data points. It addresses uncertainties arising from human annotation errors, ambiguous instances, or inherent data noise. Early pioneers such as [29] and [30] utilized counting methods to estimate false positive and false negative rates in binary classification. [31] introduced thresholding to fortify against epistemic errors in predicted probabilities and class imbalances. However, confident learning generalizes the application of thresholds to accommodate multi-class noisy labels and reweighs the loss during training to reflect the adjusted priors for the data removed.

This adaptation is informed by foundational research equating learning with noisy labels to risk minimization through loss reweighing [32,33]. More recently, [34] proposed a pragmatic deep self-supervised learning strategy that avoids probabilities by using embedding layers of a neural network. In contrast, confident learning is non-iterative and theoretically grounded. [35] estimated label noise employing methods based on confusion matrices and cross-validation, presupposing a more constrained form of label shift than class-conditional noise. [36] corroborated the empirical benefit of first rectifying label errors before training on cleaned data; however, it is limited to uniform (symmetric) and pair label noise. Confident learning augments these empirical observations and provides a theoretical justification for a broader spectrum of asymmetric and class-conditional label noise.

3 Proposed Approach

The One Shot Removal (OSR) method is an approach implemented during the training stage of machine learning models, aimed at refining the dataset through the systematic elimination of outlier data points in a single iteration. This is achieved by utilizing computed anomaly scores to identify data points that deviate significantly from the model's definition of normal samples. As illustrated in Fig. 2, the OSR method targets outlier removal as a one-time operation to enhance the dataset and improve the model's ability to detect and interpret anomalous samples present within the data. Algorithm 1 illustrates the whole procedure.

3.1 Anomaly Score Calculation

The anomaly score computation for each data point x_i is performed using a scoring function $S(x)$ affected by the outcomes of a normalizing flow model. The model, a probabilistic approach, precisely estimates the likelihood of occurrence of input data, which is particularly advantageous in an anomaly detection context. Our methodology is agnostic in terms of anomaly detection model and can be applied to any unsupervised method. Nevertheless, for our preliminary evaluation of the impact of our approach, we focused on DifferNet approach [21], a state-of-the-art SDD method that leverages the principles of normalizing flows for robust anomaly detection. A detailed description of the technique is given below:

1. In order to extract features from the input data point x_i, initially, the model employs a convolutional neural network (CNN). This transforms the raw data into a feature representation that encapsulates its essential characteristics.

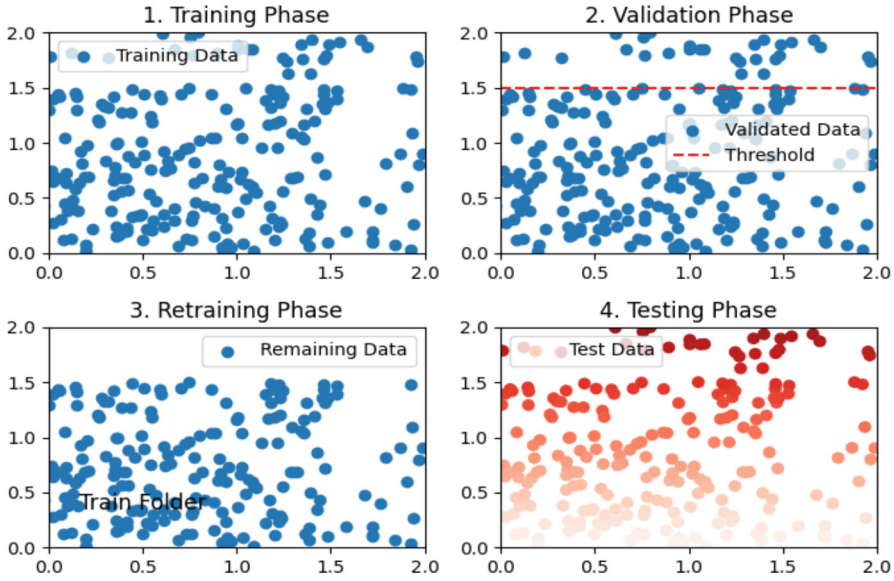

Fig. 2. The systematic procedure for improving algorithms via training, validation, retraining, and testing.

2. The retrieved characteristics are then fed into a normalizing flow model. This model estimates the probability density of the feature model in a latent space with reduced dimensions using probability density modeling. The parameters θ of the normalizing flow model, acquired during training, facilitate data conversion into a latent space, enabling density computation.
3. The anomaly score $S(x_i)$ is calculated using the likelihood given by the normalizing flow model. In our model, a lower likelihood corresponds to a higher anomaly score, indicating the likelihood that data point x_i is an anomaly. Specifically, $S(x_i) = -\log p(x_i; \theta)$, where $p(x_i; \theta)$ denotes the probability density of x_i estimated by the normalizing flow model with parametrization θ.

The model training aims to maximize the likelihood of normal data points while simultaneously minimizing the likelihood of anomalies. The model's ability to differentiate between normal and anomalous data is enhanced by the adjustment of model parameters, denoted as θ, throughout the training process. This method enables our model to detect anomalies by assessing each data point's correspondence to the trained model of normal data. The anomaly score quantifies the extent of divergence from the model, with higher scores indicating greater variance and a higher likelihood that the observed data is an anomaly.

3.2 Data Cleaning

After the training process, the next step involves removing data points that exceed a predetermined threshold value, following the calculation of anomaly scores. To exclude

the most significant outlier from the dataset, we identify the data point x_m with the maximum score $S(x)$:

$$D_{\text{new}} = D_{\text{old}} - \{x_m\} \tag{1}$$

where D_{new} represents the dataset after outlier exclusion, and D_{old} represents the dataset before outlier exclusion. The outliers with the highest anomaly score, indicating the most notable deviation from the norm, are represented by x_m.

3.3 Threshold Setting

A predetermined threshold value T is established before commencing the training process. This threshold signifies the minimum dataset size expected to remain after outlier removal. The outlier elimination process continues until the dataset size $|D|$ reaches the predetermined threshold:

$$\text{Continue removal if } |D| > T \tag{2}$$

Implementing this threshold helps mitigate the risk of excessive dataset pruning, which could lead to the loss of valuable information or the development of an overfitted model based on excessively "cleaned" data.

Once the threshold value T is reached, the model undergoes training using the refined dataset. The remaining data is considered a "normal" data distribution. Training on clean data not only enhances the model's performance on unseen data but also improves its capacity to generalize and provide accurate predictions.

Algorithm 1. Outlier Removal.

Input: D - Dataset, T - Threshold value, θ - Model parameters
Output: θ - Updated model parameters

while len$(D) > T$ **do**
  Calculate anomaly scores for the dataset
  $scores \leftarrow [S(x, \theta)$ for $x \in D]$
  Identify the data point with the higher scores
  $x_m \leftarrow \max(D,$ key=lambda $x : S(x, \theta))$
  Exclude the outliers from the dataset
  D.remove(x_m)
  Update the model parameters (θ) based on the refined dataset
  $\theta \leftarrow$ update_model(D, θ)
end while

4 Experiments

In this study, we investigate the effectiveness of our approach in training an anomaly detection model in presence of noisy labels for the training data.

4.1 Dataset

We evaluate the performance of our approach using the publically available Kolektor SDD2 dataset [2]. The dataset comprises RGB images that shows defective production items, meticulously sourced and annotated by Kolektor Group d.o.o. The defects demonstrate a wide range of attributes, including variations in size, shape, and color. These imperfections extend from small scratches and spots to substantial faults on the surface. To ensure uniformity, we standardize the dataset's resolution by center-cropping and resizing all images to dimensions of 448×448 pixels. Subsequently, the dataset is divided into several subsets for training and testing purposes. The training set comprises a total of 2085 normal samples and 246 positive samples, while the test set includes 894 negative samples and 110 positive samples.

4.2 Performance of DifferNet Model

We started our experiments by training an anomaly detection model using a dataset comprising 100 'good' images randomly sampled from the KSDD2 training set, achieving an AUROC score of 0.92 when tested on 200 random samples from the testing set, equally split between good and anomaly images. We systematically introduced anomalous images into the training dataset to simulate varying degrees of data contamination while maintaining a consistent test dataset. As we increased the proportion of anomalous images in the training dataset, with a ratio eventually reaching 50 good to 50 anomalous images, the AUROC score notably decreased to 0.74. This decline indicated a reduction in the model's capacity to discern anomalies. Subsequently, the trend persisted, with a significant drop to 0.63 observed when the training dataset consisted of 40 good and 60 anomalous images. Ultimately, with the training dataset comprising only 100 anomalous images, the AUROC plummeted to 0.07, which is coherent with the original performances of the method, simply switching the class labels. The gradual decrease in AUROC, depicted in the blue curve of Fig. 3, underscores the necessity for robust models capable of generalizing effectively to unseen data and handling various anomaly types. Hence, it is imperative to develop and deploy anomaly detection models that are resilient to data variations and effectively identify anomalies even amidst different levels of data contamination. Our approach emphasizes the importance of robustness in anomaly detection, advocating for methods capable of adapting to diverse datasets and varying levels of data quality while maintaining simplicity and effectiveness.

4.3 Performance of OSR Approach

After conducting a performance analysis of the model at different noise levels, we implemented our approach to mitigate noise-related issues. Our strategy focused on refining the training and validation processes to enhance the model's anomaly detection capabilities. Through experimentation, we carefully selected a threshold value of 1.5 to balance noise filtering and retention of relevant data for training and validation. This threshold improved the model's resilience to noise while preserving critical information essential for anomaly detection. Results demonstrated significant improvements in the model's performance across all noise levels, particularly with an AUROC score

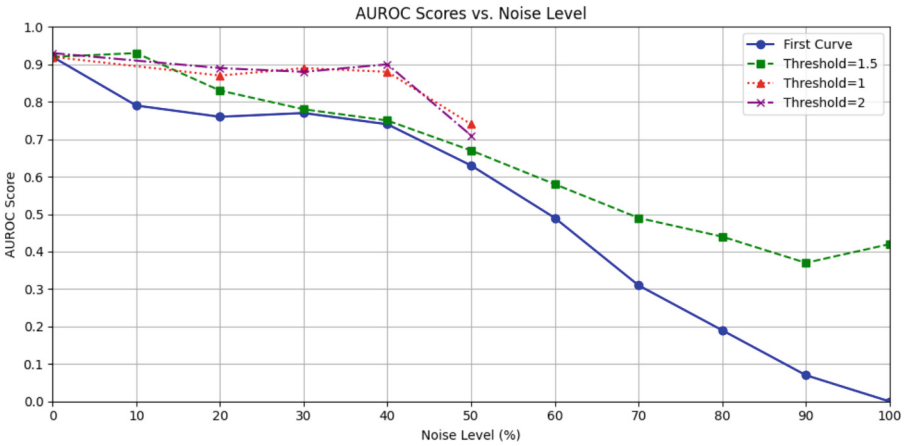

Fig. 3. The performance of the model at different noise levels before and after implementing our approach. The blue line represents the performance of the original model. The purple, red, and green curves with different threshold values represent the performance after implementing the approach. (Color figure online)

increase compared to the original model. Notably, our approach with a threshold value of 1.5 exhibited superior anomaly detection capability, as shown in green curve in Fig. 3.

In industrial settings where accuracy is paramount, model robustness is crucial. Our approach, characterized by a threshold value of 1.5, demonstrated exemplary robustness in the presence of noise, maintaining relatively stable performance across varying noise levels. With a noise level of 0 percent, our approach demonstrated an AUROC score of 0.92, suggesting its effective and precise anomaly detection capability without introducing interference. As 50 percent noise was introduced, a noticeable decrease in the AUROC score was observed, ultimately reaching 0.67. This decrease, although anticipated, highlights the negative impact of noise on the ability to distinguish patterns. However, the maintained moderate performance at this noise level suggests a degree of robustness, indicative of the model's viability under sub-optimal conditions. Upon reaching 100 percent noise, the model's AUROC score decreased to 0.42. This score is not negligible despite falling below the ideal threshold for effective anomaly detection. We hypothesize that this result is due to the ability of the model to select only a subsample of anomalies and thus increase the number of false negatives in the confusion matrix. The model's capacity to extract discernible patterns from heavily contaminated data is noteworthy, implying a level of adaptability and utility even in extreme scenarios. Compared to the original model at 100% noise, the AUROC score was 0.07, demonstrating the model's robustness by our approach.

However, experiments with threshold values of 1 and 2 for 'red' and 'purple' curves, respectively, revealed unexpected performance declines under high noise conditions, as demonstrated in Fig. 3. This emphasized the importance of robust surface detection systems in industrial settings and the need for methodologies empowering models to accurately adapt to diverse data conditions.

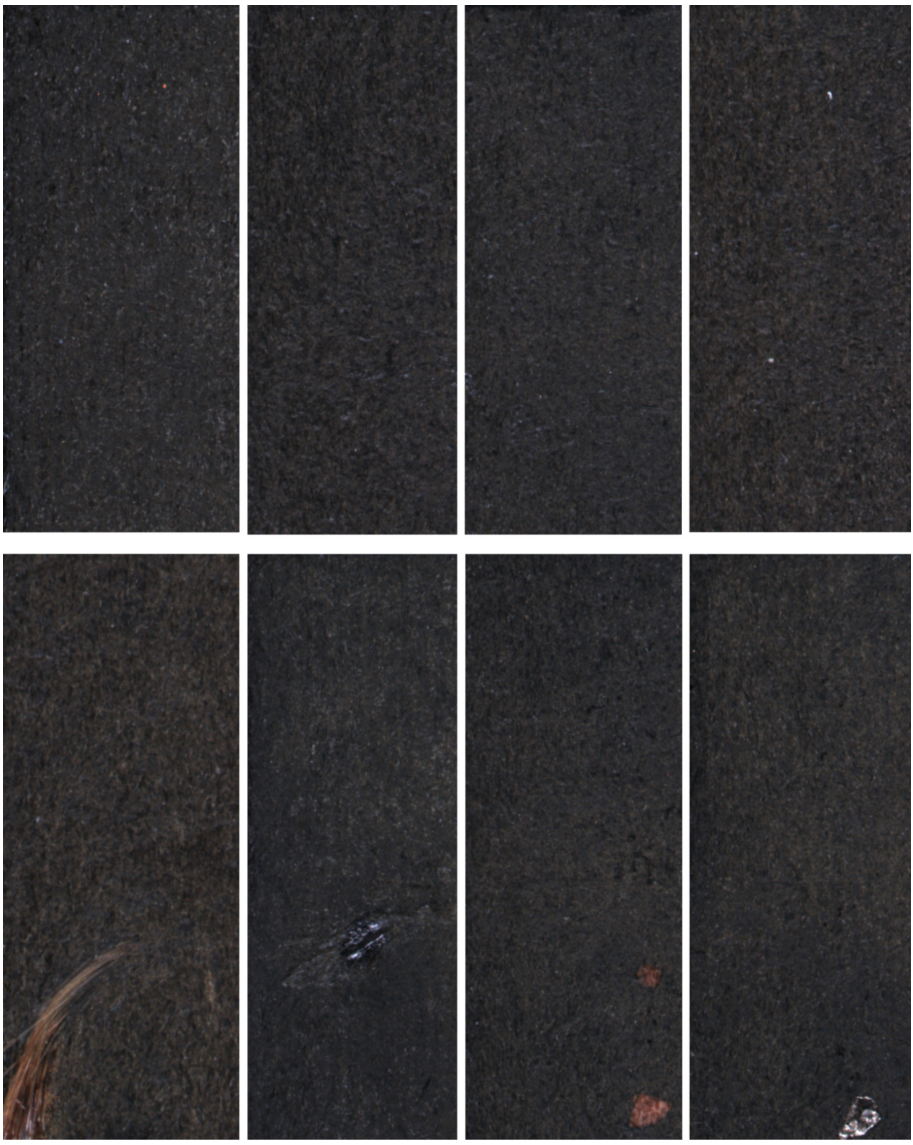

Fig. 4. The images deleted after validation exceeding threshold value which is 1.5. The top row shows good images that were deleted, while the bottom row shows anomalous images deleted.

We also report in Table 1 the number of training samples filtered out by OSR method. In case no anomalous images are provided in the training set, OSR filters out 21 images representing the less prototypical samples. It is worth noting how this filter does not generate overfitting effects since the method's overall performance is the same as the vanilla version but achieved with only 79 training samples. When anomalous

samples corrupt the dataset, OSR can filter out a majority of anomalous images. For example with 10% anomalies, OSR can reject 8 out of 10 bad samples.

Table 1. Number of Good and Anomaly Images Deleted After Validation Above Threshold Value.

Good Images	Anomaly Images	Good Del.	Anomaly Del.	Total Deleted
100	0	21	0	21
90	10	9	8	17
80	20	5	13	18
70	30	4	12	16
60	40	9	24	33
50	50	8	15	23
40	60	9	18	27
30	70	7	17	24
20	80	6	19	25
10	90	3	18	21
0	100	0	25	25

4.4 Threshold Optimization

Throughout the training and validation phases, we applied a noise threshold to filter out images exceeding the specified level. This adjustment proved pivotal in enhancing the quality of the training dataset, subsequently improving the model's capability to differentiate between normal and anomalous data. Figure 4 shows some samples of the discarded images resulting from our optimized approach, providing visual insights into the selection process of the training dataset and its impact on the model's performance.

5 Conclusions

In this study, we discussed about the usage of Self-Supervised Learning in the field of Surface Defect Detection in order to increase robustness to noise on the labels provided at training time. We have discussed the state of the art in all the related fields, and we proposed a baseline mechanism that integrates thresholding to filter uncertain data during training. Our experiments conducted on the challenging KSDD2 dataset, highlight the critical importance of developing robust anomaly detection models tailored for industrial settings. Our approach has demonstrated significant improvements in model performance through careful selection of threshold values, notably enhancing anomaly detection accuracy and reliability. These findings underscore the potential of self-supervised learning techniques for addressing robustness challenges in SDD tasks, paving the way for further advancements in industrial defect detection systems. Future research could explore additional refinements to our approach and its application in real-world manufacturing environments.

Acknowledgments. This study was carried out within the PNRR research activities of the consortium iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) - Missione 4 Componente 2, Investimento 1.5 - D.D. 1058 23/06/2022, ECS_00000043).

Disclosure of Interests. The authors declare that they have no conflict of interest.

References

1. Vrochidou, E., et al.: Towards robotic marble resin application: crack detection on marble using deep learning. Electronics **11**(20), 3289 (2022)
2. Božič, J., Tabernik, D., Skočaj, D.: Mixed supervision for surface-defect detection: from weakly to fully supervised learning. Comput. Ind. **129**, 103459 (2021)
3. Jawahar, M., Anbarasi, L.J., Geetha, S.: Vision based leather defect detection: a survey. Multimedia Tools Appl. **82**(1), 989–1015 (2023)
4. Zavrtanik, V., Kristan, M., Skočaj, D.: DSR-a dual subspace re-projection network for surface anomaly detection. In: Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T. (eds.) ECCV 2022. LNCS, vol. 13691, pp. 539–554. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-19821-2_31
5. Luo, Z., He, K., Yu, Z.: A robust unsupervised anomaly detection framework. Appl. Intell. **52**(6), 6022–6036 (2022)
6. Hu, M., Wang, Y., Feng, X., Zhou, S., Wu, Z., Qin, Y.: Robust anomaly detection for time-series data (2022)
7. Ono, Y., Tsuji, A., Abe, J., Noguchi, H., Abe, J.: Robust detection of surface anomaly using LiDAR point cloud with intensity. Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci. **43**, 1129–1136 (2020)
8. Beggel, L., Pfeiffer, M., Bischl, B.: Robust anomaly detection in images using adversarial autoencoders. In: Brefeld, U., Fromont, E., Hotho, A., Knobbe, A., Maathuis, M., Robardet, C. (eds.) ECML PKDD 2019. LNCS, vol. 11906, pp. 206–222. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46150-8_13
9. Rousseeuw, P., Perrotta, D., Riani, M., Hubert, M.: Robust monitoring of time series with application to fraud detection. Econometrics Stat. **9**, 108–121 (2019)
10. Zhao, Z., et al.: Robust anomaly detection on unreliable data. In: IEEE/IFIP International Conference on Dependable Systems and Networks (DSN) (2019)
11. Rousseeuw, P.J., Hubert, M.: Anomaly detection by robust statistics. Wiley Interdisc. Rev. Data Min. Knowl. Discov. **8**(2), e1236 (2018)
12. Zhou, C., Paffenroth, R.C.: Anomaly detection with robust deep autoencoders. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2017)
13. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2019)
14. Zhao, Z., et al.: RAD: on-line anomaly detection for highly unreliable data. arXiv preprint arXiv:1911.04383 (2019)
15. De Vitis, G.A., Foglia, P., Prete, C.A.: Row-level algorithm to improve real-time performance of glass tube defect detection in the production phase. IET Image Proc. **14**(12), 2911–2921 (2020)
16. Chen, Y., Ding, Y., Zhao, F., Zhang, E., Wu, Z., Shao, L.: Surface defect detection methods for industrial products: a review. Appl. Sci. **11**(16), 7657 (2021)

17. Bhatt, P.M., et al.: Image-based surface defect detection using deep learning: a review. J. Comput. Inf. Sci. Eng. **21**(4), 040801 (2021)
18. Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: GANomaly: semi-supervised anomaly detection via adversarial training. In: Jawahar, C., Li, H., Mori, G., Schindler, K. (eds.) ACCV 2018. LNCS, vol. 11363, pp. 622–637. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-20893-6_39
19. 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 (ICPR) (2021)
20. Roth, K., Pemula, L., Zepeda, J., Schölkopf, B., Brox, T., Gehler, P.: Towards total recall in industrial anomaly detection. In: IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2022)
21. Rudolph, M., Wandt, B., Rosenhahn, B.: Same same but differnet: semi-supervised defect detection with normalizing flows. In: IEEE/CVF Winter Conference on Applications of Computer Vision (2021)
22. Zavrtanik, V., Kristan, M., Skočaj, D.: Draem-a discriminatively trained reconstruction embedding for surface anomaly detection. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2021)
23. Capogrosso, L., et al.: Diffusion-based image generation for in-distribution data augmentation in surface defect detection. In: International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, vol. 2, pp. 409–416. SciTePress (2024)
24. Božič, J., Tabernik, D., Skočaj, D.: End-to-end training of a two-stage neural network for defect detection. In: International Conference on Pattern Recognition (ICPR). IEEE (2021)
25. Zhang, C., Wang, Z., Liu, B., Xiaolei, W., et al.: Steel plate defect recognition of deep neural network recognition based on space-time constraints. Adv. Multimedia **2022**, 9595286 (2022)
26. Masci, J., Meier, U., Ciresan, D., Schmidhuber, J., Fricout, G.: Steel defect classification with max-pooling convolutional neural networks. In: International Joint Conference on Neural Networks (IJCNN) (2012)
27. Tian, R., Jia, M.: DCC-CenterNet: a rapid detection method for steel surface defects. Measurement **187**, 110211 (2022)
28. Zhang, S., Zhang, Q., Gu, J., Su, L., Li, K., Pecht, M.: Visual inspection of steel surface defects based on domain adaptation and adaptive convolutional neural network. Mech. Syst. Sig. Process. **153**, 107541 (2021)
29. Elkan, C.: The foundations of cost-sensitive learning. In: International Joint Conference on Artificial Intelligence (IJCAI), vol. 17, pp. 973–978. Lawrence Erlbaum Associates Ltd. (2001)
30. Forman, G.: Counting positives accurately despite inaccurate classification. In: Gama, J., Camacho, R., Brazdil, P.B., Jorge, A.M., Torgo, L. (eds.) ECML 2005. LNCS, vol. 3720, pp. 564–575. Springer, Heidelberg (2005). https://doi.org/10.1007/11564096_55
31. Elkan, C., Noto, K.: Learning classifiers from only positive and unlabeled data. In: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2008)
32. Natarajan, N., Dhillon, I.S., Ravikumar, P.K., Tewari, A.: Learning with noisy labels. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 26 (2013)
33. Van Rooyen, B., Menon, A., Williamson, R.C.: Learning with symmetric label noise: the importance of being unhinged. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 28 (2015)
34. Han, J., Luo, P., Wang, X.: Deep self-learning from noisy labels. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2019)

35. Lipton, Z., Wang, Y.X., Smola, A.: Detecting and correcting for label shift with black box predictors. In: International Conference on Machine Learning (ICML) (2018)
36. Huang, J., Qu, L., Jia, R., Zhao, B.: O2U-Net: a simple noisy label detection approach for deep neural networks. In: IEEE/CVF International Conference on Computer Vision (ICCV) (2019)