



Cutout as augmentation in contrastive learning for detecting burn marks in plastic granules

Muen Jin and Michael Heizmann

Institute of Industrial Information Technology, Karlsruhe Institute of Technology, 76131 Karlsruhe, Germany

Correspondence: Muen Jin (muen.jin@kit.edu)

Received: 30 September 2023 – Revised: 8 February 2024 – Accepted: 20 February 2024 – Published: 4 April 2024

Abstract. Plastic granules are a common delivery form for creating products in industries such as the plastic manufacturing, construction and automotive ones. In the corresponding sorting process of plastic granules, diverse defect types could appear. Burn marks, which potentially lead to weakened structural integrity of the plastic, are one of the most common types. Thus, plastic granules with burn marks should be filtered out during the sorting process. Artificial intelligence (AI)-based anomaly detection approaches are widely used in the field of visual-based sorting due to the higher accuracy and lower requirement of expert knowledge compared with classic rule-based algorithms (Chandola et al., 2009). In this contribution, a simple data augmentation strategy, *cutout*, is implemented as a way of simulating defects when combined with a contrastive learning-based methodology and is proven to improve the accuracy of the anomaly detection of burn marks. Different variants of cutout are also evaluated. Specifically, synthetic image data are used due to the lack of real data.

1 Introduction

Plastic granules used in industry have various defect types (Peršak et al., 2021). A common defect, known as a burn mark, arises from either overheating during plastic injection or mechanical stress causing excessive friction or pressure during processing. Burn marks can be observed in the form of dark spots or whole discolored granule surfaces. Burn marks are not merely a visual defect. Moreover, they indicate the degradation of both physical and chemical properties of the corresponding parts compared with the intact parts. Plastic granules with burn marks should be identified and filtered out by a sorting system. In Peršak et al. (2021), a color feature-based classifier was employed to sort various defect types of granules, including burn marks.

To incorporate machine learning methods for increasing the detection accuracy of defective granules, sufficiently well-labeled data are required. Most defective granules take up only a fraction of the total amount, including those with burn marks, which increases the cost of labeling for these granules. Meanwhile, normal granules are easily accessible in practice, e.g., in a defect-free running period of the production. Normal plastic granule images can be obtained using methods like threshold-based foreground–background seg-

mentation, blob detection, or neural-network-based segmentation, depending on the scene's complexity. Based on this assumption, the sorting task is regarded as a one-class classification problem in this work; i.e., only images of normal granules are available during training.

Instead of using real-world data without reliable ground-truth labels, synthetic data of plastic granules generated by a blender were used in this work. Multiple granule instances were modeled in each rendering step to simulate the real-world captured image from line cameras over the conveyor belt in the sorting system. The synthetic data only contain normal granules and granules with burn marks; other defect types are not modeled. One advantage of using synthetic data is that both the precise location and the corresponding ground-truth label of each granule in the rendered image are accessible without extra effort. Each plastic granule was cropped from the rendered image and labeled for further processing, such as classification or anomaly detection (Fig. 1).

As a large number of normal data are available during training, some of the synthetic images of normal granules can be leveraged for pre-training in a self-supervised manner using a contrastive learning method, since contrastive representation shows state-of-the-art performance in visual recog-

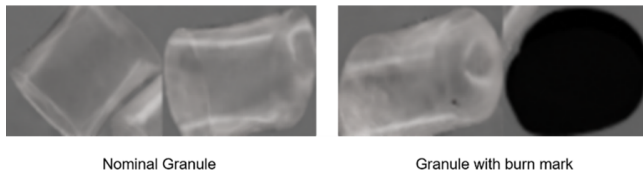


Figure 1. Synthetic images of granules generated using a blender.

tion tasks (Chen et al., 2020; van den Oord et al., 2019; He et al., 2020). We followed the work of Sohn et al. (2020). By introducing cutout (DeVries and Taylor, 2017) as an extra augmentation technique in this framework, the detection accuracy of granules with burn marks increased. We show that the increase is largely attributed to the visual similarity between the burn marks and the cutout masks.

2 Related works

Synthetic data are increasingly utilized to improve model performance in classification and detection tasks. Synthetic data are generated through domain randomization to train deep neural networks for robotic grasping tasks, improving the model's generalization to real-world scenarios (Sankaranarayanan et al., 2017). Semantic segmentation performance has been enhanced by integrating synthetic data, effectively augmenting the original dataset and improving the model robustness (Zhu et al., 2018).

Many algorithms for anomaly detection are based on learning features from nominal data. This is especially the case for industrial anomaly detection for bulk materials, as collecting defect samples can be costly, while nominal samples are available in a sufficient amount for training feature-learning-based algorithms. Anomaly detection is regarded as one-class classification in scenarios where the objective is to identify and classify outliers as a single distinct class and only data of nominal samples are available during training. One-class classification often grapples with the lack of detailed semantic information. Various techniques, such as leveraging auto-encoder (Sakurada and Yairi, 2014) or Gaussian mixture models (Zong et al., 2018) or even attention-guided methods (Venkataraman et al., 2020), have been proposed to refine the estimation of nominal feature distributions. Generative models capable of capturing data density have been employed to discern outliers through low-density samples (Zong et al., 2018; Schlegl et al., 2017). However, interpreting density in high-dimensional spaces remains a challenge. Self-supervised learning, a common strategy for learning representations from unlabeled data (Noroozi and Favaro, 2016; Caron et al., 2018), has also been extended to one-class classification. For instance, contrastive learning has been employed to enhance out-of-distribution detection within a multi-class setting (Winkens et al., 2020). Self-supervised contrastive learning learns representations

by training models to distinguish data applied with different augmentations (e.g., color jittering) from other data instances. In Sohn et al. (2020), a two-stage framework focusing on learning from a singular class of examples and proposing innovative distribution-augmented contrastive learning was proposed. In the first stage, a deep neural network is trained using self-supervised contrastive learning for obtaining a high-level data representation. In the second stage, a single one-class classification such as kernel density estimation (KDE) or one-class support vector machine is built using the representation from the first stage. Instead of using surrogate losses (Golan and El-Yaniv, 2018; Hendrycks et al., 2019), building a separate one-class classifier is proven to be more faithful to the classification task and hence showcases consistent performance improvements. The representation learned using contrast has two problems when used for the one-class classifier. The first problem is the class collision, as in contrastive learning the distance between representations of different samples should be maximized, even when they are of the same class. The common practice of one-class classification however minimizes the distance between representations with respect to the class center. The second problem is the uniformity of representations. It is proven that a perfectly trained model using contrastive learning will project all data onto the surface of a hypersphere uniformly, which potentially makes the abnormal data at the inference time less distinguishable (Ruff et al., 2018). To overcome these two problems, the distribution augmentation is introduced in the first stage. Except for the normal data augmentations, geometric transformations including rotation and flip are applied to data instances. The distribution augmentation is disjoint from data augmentation. An image and its distribution-augmented version are considered two separate instances and hence are encouraged to be distant in the representation space. Essentially, by introducing distribution augmentation, the number of data instances for training is increased. Furthermore, the uniformity of representations in the hypersphere is eased. Another way of learning representation in a self-supervised manner is augmentation prediction (Gidaris et al., 2018). By discriminating augmentation applied to data (e.g., the rotation degree of the image), the model learns deep representations which approximate the normality score of data used for one-class classification well. Similarly, Li et al. (2021) introduce *cutpaste* as an augmentation technique, which cuts a small rectangular area from the normal image and then pastes it back into the image at a random location. A model is pre-trained to identify whether an image is applied with cutpaste. This self-supervised pre-training method is proven to be beneficial for downstream defect detection and localization tasks.

This work is mainly based on the two-stage framework in Sohn et al. (2020). The only modification is the choice of distribution augmentation. Instead of using geometric transforms, cutout is applied to increase the model's sensitivity for detecting burn marks. Cutout involves setting pixel values of

a random rectangular or square area in the image to zero. It was originally introduced as an augmentation technique to train classification models by encouraging models to learn robust and discriminative features which are invariant to local changes. In this work, cutout masks are applied as distribution augmentation to mimic the burn marks. Since the data instances with and without distribution augmentation are implicitly labeled during contrastive learning, the pre-trained model is adapted to distinguish between nominal granules and granules with burn marks in the second stage.

3 Experiments

3.1 Experimental details

Dataset. To study the detection performance of plastic granules with burn marks, our experiments are performed on the synthetic grayscale image data of granules generated using the blender. The synthetic dataset for training contains 4214 images of normal granules, and the dataset for the test contains 1205 images of normal granules and 341 images of granules with burn marks. Images are resized to 256×256 and no center crop is applied, as the images are cropped from the rendering based on a minimal unoriented bounding box. During the contrastive learning phase, normal data augmentations including crop and resize, flip, random grayscale and blur are applied.

Distribution augmentation. Except for the rotation and flip used in Sohn et al. (2020), cutout is applied as the distribution augmentation. Concretely, each image will be applied with a rectangular cutout mask at a random location which set the pixel values within the area to zero before the necessary normalization step. The size of the cutout mask is chosen randomly in the range from 2% to 80% of the image size. The aspect ratio of the cutout mask is also chosen randomly. We choose the rectangular mask instead of the default square mask due to the variance of the real burn mark shapes. We also tested different variants of the cutout masks, and the results are shown in Sect. 3.2.

Model and parameter setting. As in Sohn et al. (2020), Resnet-18 is used as a feature extractor. We use the momentum stochastic gradient descent as an optimizer to pre-train the model with 200 epochs with a cycle cosine learning rate decay for the contrastive learning of the Resnet-18 model. A kernel density estimation detector is applied in the second stage for classification.

Metric. Following Sohn et al. (2020), detection accuracy is measured via the area under the receiver operator curve (AUROC). Only image-level anomaly detection performance is measured.

3.2 Experimental results

The result of image-level burn mark detection is shown in Table 1. The AUROC is averaged over five runs for each pre-

Table 1. Results of burn mark detection. The bold font means the methodology achieves the best results of all.

Pre-training methodology	AUROC (%)
Contrastive	80.7 ± 1.6
Contrastive + rotation and flip (DA)	86.7 ± 1.2
Contrastive + cutout (DA)	90.7 ± 0.3
Binary classification + cutout	85.2 ± 0.9

training methodology. Table 1 reveals that applying distribution augmentation increases the image-level detection accuracy of burn marks, and replacing rotation with cutout as the distribution augmentation further increases the accuracy by 4%. Meanwhile, using cutout as a distribution augmentation leads to smaller variance, which suggests more robust pre-trained network weights. In addition, we also tested the simpler pre-training for a binary classification following Li et al. (2021) but using cutout instead of cutpaste. Compared with contrastive learning, no data augmentation is applied, and Resnet-18 is pre-trained to determine whether an image was applied with cutout or not. The detection accuracy is better than contrastive learning without distribution augmentation but worse than when rotation and flip or cutout is applied as a distribution augmentation.

In Fig. 2, the features of image data extracted using the contrastive learning pre-trained network are visualized using t-distributed stochastic neighbor embedding (t-SNE). The images of granules with burn marks and the images applied with cutout masks share similar statistics in the two-dimensional space. Some of both kinds of data lie in the cluster of normal granules, and the rest form subclusters close to each other. Such statistics imply that the images applied with cutout masks do not completely mimic the appearance of the granules with real burn marks, yet the burn marks and cutout masks share similar semantic features in feature space. As a result, applying cutout as the distribution augmentation during contrastive learning introduces knowledge of burn marks to the model initialization without access to images of granules with real burn marks.

Variants of cutout masks. We investigate the impact of different cutout mask variations, considering factors such as size, orientation, shape or grayscale. The results of this study are presented in Table 2. In the case of a fixed mask size, the mask is chosen as either 2% or 80% of the image size with 50% probability each, which is also the minimal or maximal size in the default setting. These two values are chosen based on the sizes of real burn marks. All these variants lead to slightly worse detection performance compared to the default setting. Figure 3 shows the features of images extracted after using fixed-size cutout masks for contrastive learning. The subclusters of granules with cutout masks and granules with burn marks on the right are more concentrated and less overlapped, implying a smaller similarity between them. Fig-

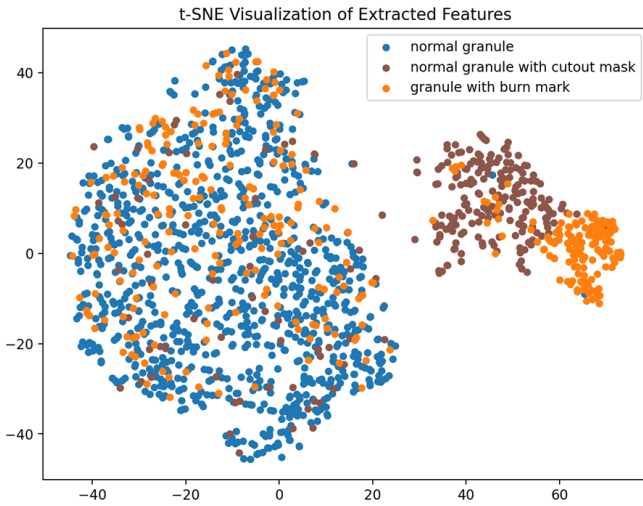


Figure 2. Visualization of representations pre-trained with distribution-augmented contrastive learning.

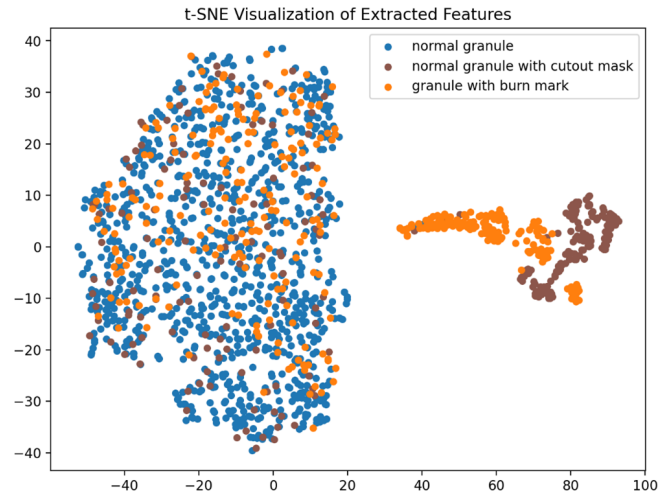


Figure 3. Visualization of representations after pre-training using a cutout mask of a fixed size.

Table 2. Influence of the different cutout masks. The bold font means the methodology achieves the best results of all.

Cutout mask	AUROC (%)
Default (2%–80%)	90.7
Fixed size (2%/80%)	87.3
Default + rotation	89.3
Circle	88.0
Grayscale 64	88.3
Grayscale 128	88.1
Grayscale 256	90.4

ures 4, 5 and 6 show results of applying cutout masks with different grayscales. In Figs. 4 and 6, in comparison to the default mask with a grayscale value of 0, the subclusters of images applied with gray or white cutout masks are further from images containing real burn marks. In Fig. 5, the images applied with cutout masks are almost inseparable from the main cluster of normal granules without burn marks, as in this scenario the cutout masks closely resemble the image background. This result illustrates in another aspect the similarity between black cutout masks and real burn marks.

Lastly, we tested the classification accuracy of the trained model on 200 images of real granules, while 37 of them have burn marks. The confusion matrix of the result is shown in Table 3. The results in parentheses are achieved by a model that applies the default rotation as the distribution augmentation during training, while those outside the parentheses are achieved by a model that applies the cutout as the distribution augmentation. The F1 score increases from 0.9268 to 0.9606.

Due to the limited number of real test data, we further tested both models on an extra synthetic test dataset which contains 10 000 images, while 1850 of them have burn marks. The proportion of images containing burn marks is

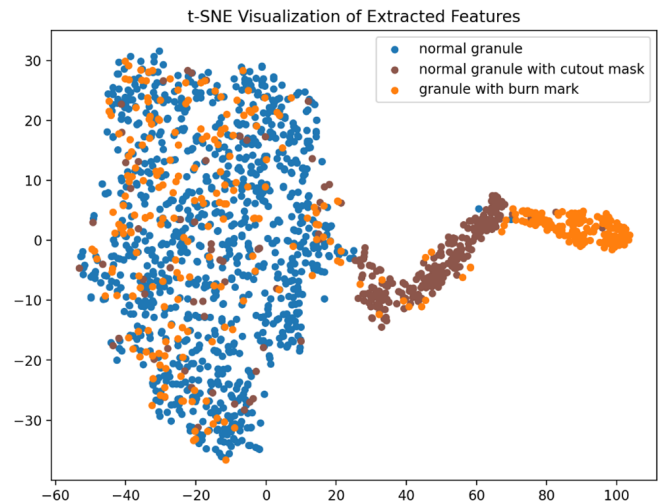


Figure 4. Visualization of representations pre-trained using a cutout mask of grayscale 64.

the same as that of the real test data. We use the Fréchet inception distance (FID) for measuring the similarity between the real test data and the synthetic test data. A lower FID score represents a higher similarity. The FID score between the real data and our synthetic data is 74.8, which is low enough compared to another recent result of industrial image generation (Zhong et al., 2023), and hence the similarity between the real test data and the synthetic data can be validated. We tested the same models with the corresponding thresholds again on the extra synthetic test data and the results shown in Table 4. Compared to the results on real test data, similar improvements in the true positive rate and the true negative rate can be observed. Meanwhile, both the false negative rate and the false positive rate drop when applying cutout instead of the default rotation as a distribution aug-

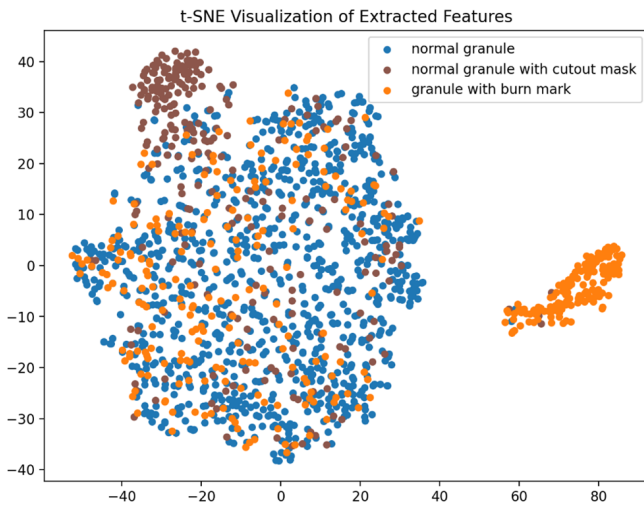


Figure 5. Visualization of representations pre-trained using a cutout mask of grayscale 128.

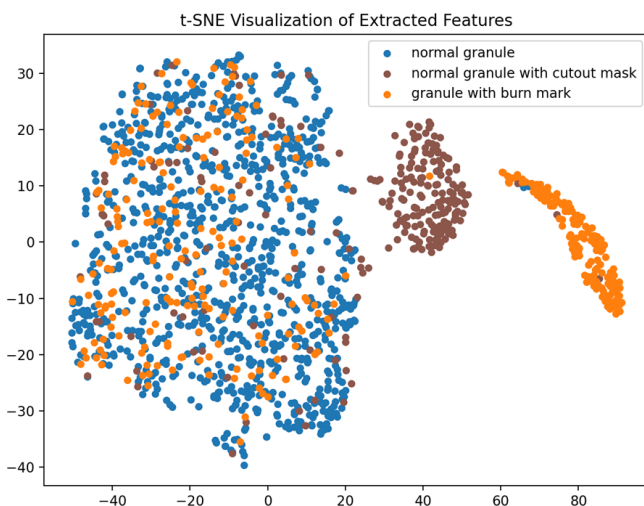


Figure 6. Visualization of representations pre-trained using a cutout mask of grayscale 255.

mentation for contrastive learning. The F1 score increases from 0.9292 to 0.9643. This improvement is consistent with that from the real test data. Besides, the overall performance on the synthetic test data is better than that on real test data, which is reasonable since the domain gap between the synthetic test data and the synthetic data used for training is smaller compared to the real data.

4 Conclusion and outlook

In this study, we have demonstrated the efficacy of employing the data augmentation strategy cutout in combination with a contrastive learning-based methodology to enhance the accuracy of burn mark detection in plastic granules. The experimental results highlight the substantial improvement

Table 3. Confusion matrix of the classification result on real images.

		Actual	
		normal	defect
Predicted	normal	159 (152)	9 (13)
	defect	4 (11)	28 (24)

Table 4. Confusion matrix of the classification result on synthetic test images.

		Actual	
		normal	defect
Predicted	normal	8011 (7623)	454 (634)
	defect	139 (527)	1396 (1216)

achieved through this approach, showcasing its potential to refine anomaly detection processes in industrial settings. The utilization of synthetic data generated by blender-facilitated efficient training addressed the challenge of limited well-labeled real-world data. The augmentation technique proved particularly effective at enhancing the model's sensitivity in detecting burn marks, contributing to a more robust and accurate detection system. Furthermore, the improvements of burn mark detection accuracy brought by introducing cutout reveals the potential of defect simulation using simple augmentation without extensive prior knowledge of the visual characteristics of the defect type. For example, for those granules with a blurred surface (Peršak et al., 2021), which are also referred to as cloudy plastic granules, a partial image blurring can be similarly applied as a distribution augmentation method for contrastive learning. This is however outside the scope of this work.

While the results are promising, certain limitations should be acknowledged. The study focused on a specific defect type, i.e., burn marks, and utilized synthetic data representing a simplified scenario. Future research should encompass a broader spectrum of defect types and incorporate real-world data to create a more comprehensive and diverse dataset. Additionally, investigating the applicability of this methodology to a wider array of industrial contexts and optimizing the size and shape of the cutout masks for various defect types would enhance its versatility and practicality. Further efforts should be directed towards addressing the challenges posed by overlapping defect features and exploring ensemble models to boost detection accuracy and reliability in complex industrial environments.

Code and data availability. The synthetic dataset can be downloaded using the following link: <https://drive.google.com/drive/folders/190ywN1Yi-C18Nmji0ZoOmZuyKtToF8zy> (Jin, 2023).

The code is not publicly accessible but can be requested from the corresponding author.

Author contributions. MJ generated the data and developed the methodology as well as the experimental setup. MH provided supervision and is the administrative support.

Competing interests. The contact author has declared that neither of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Special issue statement. This article is part of the special issue "Sensors and Measurement Science International SMSI 2023". It is a result of the 2023 Sensor and Measurement Science International (SMSI) Conference, Nuremberg, Germany, 8–11 May 2023.

Financial support. The article processing charges for this open-access publication were covered by the Karlsruhe Institute of Technology (KIT).

Review statement. This paper was edited by Thomas Fröhlich and reviewed by two anonymous referees.

References

- Caron, M., Bojanowski, P., Joulin, A., and Douze, M.: Deep clustering for unsupervised learning of visual features, in: Proceedings of the European conference on computer vision (ECCV), 8 September 2018, Munich, Germany, Springer, Cham, 132–149, https://doi.org/10.1007/978-3-030-01264-9_9, 2018.
- Chandola, V., Banerjee, A., and Kumar, V.: Anomaly detection: A survey[J]. ACM computing surveys (CSUR), <https://doi.org/10.1145/1541880.1541882>, 2009.
- Chen, T., Kornblith, S., Norouzi, M., and Hinton, G.: A simple framework for contrastive learning of visual representations, in: International conference on machine learning, PMLR, 12 July 2020, Vienna, Austria, 1597–1607, arXiv [preprint], <https://doi.org/10.48550/arXiv.2011.02578>, 4 November 2020.
- DeVries, T. and Taylor, G. W.: Improved regularization of convolutional neural networks with cutout, arXiv [preprint], <https://doi.org/10.48550/arXiv.1708.04552>, 15 August 2017.
- Gidaris, S., Singh, P., and Komodakis, N.: Unsupervised representation learning by predicting image rotations, arXiv [preprint], <https://doi.org/10.48550/arXiv.1803.07728>, 21 March 2018.
- Golan, I. and El-Yaniv, R.: Deep anomaly detection using geometric transformations, arXiv [preprint], <https://doi.org/10.48550/arXiv.1805.10917>, 28 May 2018.
- He, K., Fan, H., Wu, Y., Xie, S., and Girshick, R.: Momentum contrast for unsupervised visual representation learning, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 14 June 2020, Seattle, USA, 9729–9738, arXiv [preprint], <https://doi.org/10.48550/arXiv.1911.05722>, 23 March 2020.
- Hendrycks, D., Mazeika, M., Kadavath, S., and Song, D.: Using self-supervised learning can improve model robustness and uncertainty, arXiv [preprint], <https://doi.org/10.48550/arXiv.1906.12340>, 28 June 2019.
- Jin, M.: Synthetic dataset of plastic granules, Google Drive [data set], <https://drive.google.com/drive/folders/190ywN1Yi-C18Nmji0ZoOmZuyKrToF8zy> (last access: 24 September 2023), 2023.
- Li, C. L., Sohn, K., Yoon, J., and Pfister, T.: Cutpaste: Self-supervised learning for anomaly detection and localization, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 20–25 June 2021, Nashville, USA, IEEE, 9664–9674, <https://doi.org/10.1109/CVPR46437.2021.00954>, 2021.
- Noroozi, M. and Favaro, P.: Unsupervised learning of visual representations by solving jigsaw puzzles, in: European conference on computer vision, Springer International Publishing, Cham, 69–84, https://doi.org/10.1007/978-3-319-46466-4_5, 2016.
- Peršak, T., Viltužnik, B., Hernalds, J., and Klančnik, S.: Vision-based sorting systems for transparent plastic granulate, Appl. Sci., 10, 4269, <https://doi.org/10.3390/app10124269>, 2021.
- Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., Müller, E., and Kloft, M.: Deep one-class classification, in: International conference on machine learning, 10 July 2018, Stockholm, Sweden, PMLR, ISBN 9781510867963, 4393–4402, 2018.
- Sakurada, M. and Yairi, T.: Anomaly detection using autoencoders with nonlinear dimensionality reduction, in: Proceedings of the MLSDA 2014 2nd workshop on machine learning for sensory data analysis, 2 December 2014, Gold Coast, QLD, Australia, Association for Computing Machinery, 4–11, <https://doi.org/10.1145/2689746.2689747>, 2014.
- Sankaranarayanan, S., Balaji, Y., Jain, A., Lim, S. N., and Chellappa, R.: Learning from synthetic data: Addressing domain shift for semantic segmentation, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 21 July 2017, Honolulu, HI, USA, 3752–3761, arXiv [preprint], <https://doi.org/10.48550/arXiv.1711.06969>, 19 November 2017.
- Schlegl, T., Seeböck, P., Waldstein, S. M., Schmidt-Erfurth, U., and Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery, in: International conference on information processing in medical imaging, 25 June 2017, Boone, NC, USA, Springer International Publishing, Cham, 146–157, https://doi.org/10.1007/978-3-319-59050-9_12, 2017.
- Sohn, K., Li, C. L., Yoon, J., Jin, M., and Pfister, T.: Learning and evaluating representations for deep one-class classification, arXiv [preprint], <https://doi.org/10.48550/arXiv.2011.02578>, 4 November 2020.

- van den Oord, A. V. D., Li, Y., and Vinyals, O.: Representation learning with contrastive predictive coding, arXiv [preprint], <https://doi.org/10.48550/arXiv.1807.03748>, 22 January 2019.
- Venkataramanan, S., Peng, K. C., Singh, R. V., and Mahalanobis, A.: Attention guided anomaly localization in images, in: European Conference on Computer Vision, 23 August 2020, Glasgow, United Kingdom, Springer International Publishing, Cham, 485–503, https://doi.org/10.1007/978-3-030-58520-4_29, 2020.
- Winkens, J., Bunel, R., Roy, A. G., Stanforth, R., Natarajan, V., Ledsam, J. R., MacWilliams, P., Kohli, P., Karthikesalingam, A., Kohl, S., Cemgil, T., Eslami, S., and Ronneberger, O.: Contrastive training for improved out-of-distribution detection, arXiv [preprint], <https://doi.org/10.48550/arXiv.2007.05566>, 10 July 2020.
- Zhong, X., Zhu, J., Liu, W., Hu, C., Deng, Y., and Wu, Z.: An Overview of Image Generation of Industrial Surface Defects, *Sensors*, 23, 8160, <https://doi.org/10.3390/s23198160>, 2023.
- Zhu, X., Zhou, H., Yang, C., Shi, J., and Lin, D.: Penalizing top performers: Conservative loss for semantic segmentation adaptation, in: Proceedings of the European Conference on Computer Vision (ECCV), 8 September 2018, Munich, Germany, 568–583, arXiv [preprint], <https://doi.org/10.48550/arXiv.1809.00903>, 4 September 2018.
- Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., and Chen, H.: Deep autoencoding gaussian mixture model for unsupervised anomaly detection, in: International Conference on Learning Representations, 30 April 2018, Vancouver, Canada, ISBN 9781713872726, 2018.