Unsupervised Anomaly Detection on Volumetric data for Industrial Visual Inspection

Cynthia I. Ugwu[1,2], Sofia Casarin[1], Marco Boschetti[2] and Oswald Lanz[1,2]

[1]Free University of Bozen-Bolzano, Piazza Domenicani 3, Bolzano BZ, 39100, Italy
[2]Covision Lab, Via Julius Durst 4, Bressanone BZ, 39042, Italy

Abstract

Anomaly detection is a long-standing field of research that aims to identify anomalous patterns that differ from those seen in regular instances. In defect detection, the normal and the abnormal samples differ in their local appearance but are semantically identical, for example, defects in printed circuits, cables, or medicinal pills. There are many datasets for unsupervised defect detection at the image level which have led to the development of several methods, but some anomalies appear in the geometry or density-based property of an object which means we need a 3D approach. Although most companies already have advanced vision systems capable of capturing 2D images and 3D measurements of objects, there is a lack of 3D datasets specifically designed for defect detection and anomaly localization in industrial environments. As it is clear that 3D anomaly detection is a field that needs more exploration, we decided to focus our research on defect detection in volumetric industrial data. To achieve this goal, we first worked on the segmentation of volumetric medical data, due to a large amount of publicly available datasets and the similarities in the design principles of the architectures used for both anomaly detection and segmentation. Our final model achieved comparable results to state-of-the-art methods in the medical field by being on average ×2 faster with less than 1/3 of parameters. In the next work, we will adapt the obtained model for defect detection using our internal industrial dataset.

Keywords

Anomaly detection, Defect detection, Unsupervised learning, Volumetric data, Computer Vision

1. Introduction

Anomaly detection is a long-standing field of research with early exploration dating back to the 1960s [1]. It aims at identifying anomalous patterns that are different from those seen in regular instances. In computer vision "Anomaly Detection" has many facets, which is why the term summarizes different tasks in the literature. In the case of multi-class classification, the term is often used to describe Out-of-Distribution Detection (ODD) or novelty detection, where the task is to determine at inference time if a test sample belongs to one of the classes the model was trained with. Aside from ODD detection, one can consider two anomaly detection variants: (i) semantic anomaly detection, in which the normal and the abnormal samples differ in their semantic meaning; (ii) defect detection, in which the normal and the abnormal samples differ in their local appearance (*i.e.*, defect), but are semantically identical.

Due to the nature of the problem, applying supervised learning methods have huge drawbacks induced by the difficulty of defining and collecting enough abnormal data, the inability of the trained model to detect new types of rare events, and expensive labelling. This resulted in a methodological shift towards unsupervised or semi-supervised learning. In the field of anomaly detection, the terms unsupervised and semi-supervised are interchanged. Unsupervised learning means using data without labels, *i.e.*, using both normal and abnormal data undistinguished like in [2]. With semi-supervised learning instead, we are using only normal data. In the literature, most of the time unsupervised learning refers to using only normal data without any form of labelling (like in [3, 4, 5]).

2. Defect detection in the industrial field

It is not uncommon to see operators discriminating between good and defective parts right next to the production line. These human experts are valuable to the company and sometimes unique: since the market demand decides the number of manufactured pieces, managing work shifts is not always straightforward. Moreover, even when setting up automated systems becomes compelling, the latter's need for labelled data makes human knowledge essential. While labelling is expensive and time-consuming, it is easy to acquire data. Given the vast amount of unlabeled data available, this situation perfectly blends with the task of unsupervised defect detection and localization. In defect detection, the normal

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy
*Corresponding author.
cugwu@unibz.it (C. I. Ugwu); scasarin@unibz.it (S. Casarin); marco.moschetti@covisionlab.com (M. Boschetti); oswald.lanz@unibz.it (O. Lanz)
0000-0003-0465-6982 (C. I. Ugwu); 0000-0001-9302-3460 (S. Casarin); 0000-0003-4793-4276 (O. Lanz)

and the abnormal samples differ in local appearance but are semantically identical, for example, defects in printed circuits, cables, or medicinal pills. Usually, it is combined with defect localization that provides a heatmap indicating the location of an outlier.

MVTec AD [6] is the most widely used dataset for anomaly detection and localization. There is a glut of literature relating to image-level defect detection in the MVTec AD dataset with approaches reaching on average 99% accuracy in terms of Area Under the Receiver Operating Curve (AUROC). There is little room for improvement. The question that naturally arises is: where more research is needed?

3. 3D defect detection

Some defects manifest as anomalies in the geometric structure or density-based property of an object, which leads to the necessity of a 3D representation, *e.g.*, 3D printing, structured light and computer tomography. In recent years industrial Computed Tomography (CT) has become increasingly popular in industries. It is a non-destructive testing method for the precise examination of components and it can be used to create precise internal views of parts, weld seams, and electronic components. Through CT scans 3D information is provided by stacking multiple grayscale images to form a dense voxel grid. The final high-resolution and three-dimensional image can be used to localise and evaluate defects such as pores, voids and inclusions. Valuable for quality control is also 3D printing, where layers in the additive printing process can be imaged providing a volume of data to be analyzed for anomalies.

There is a lack of a comprehensive public 3D dataset designed explicitly for the detection and localization of anomalies. This led Bergmann *et al* to develop MVTec 3D-AD [7] dataset. In MVTec 3D-AD, the nature of data is fundamentally different from the volumetric information provided by CT scans. The dataset describes the geometric surface of objects by acquiring data that also has depth information with respect to the local camera coordinate, that is to say, we are dealing with point clouds. The best approach so far [8] reaches a detection accuracy in terms of AUROC of 72.7%. However, the authors developed a model based on strong pre-processing and handcrafted orientation-invariant representations. Other approaches apply networks that were originally developed for segmentation on medical CT scans (like [9, 10]). Unfortunately, we are still missing public datasets of volumetric scans in industrial production processes for quality control.

4. Model complexity when dealing with volumetric data

When dealing with volumetric data conventional 2D Convolutional Neural Networks (CNNs) are computationally cheap but cannot capture three-dimensional features. On the other side, although 3D CNNs are designed to learn three-dimensional features, they require higher computation costs, resulting in higher inference latency compared to 2D CNNs. Besides, the large number of parameters of 3D CNNs may result in a higher risk of overfitting, especially when encountering small datasets for training. This is very common in the medical or industrial fields as it is especially difficult to collect volumetric datasets due to accessibility issues for ethical or privacy reasons, and limited time and budget for annotations.

There have been many efforts to trade off model performance and computational complexity in other computer vision fields such as video analysis and action recognition ([11], [12], [13], [14]). It makes sense to learn spatiotemporal features using 3D convolution since a video can be seen as a temporally dense sequence of images. However, as previously mentioned, dealing with 3D CNNs is computationally intensive

5. Our model

Following the observation of a gap in the literature regarding 3D anomaly detection, we decided to focus on 3D defect detection. The largest part of our work done so far has focused on segmenting volumetric medical data, due to the big amount of publicly available datasets (like [15, 16, 17, 18]) and the similarity of architectures to the ones applicable to anomaly detection task. Both in 3D segmentation and video action recognition, there are approaches directly using 2D CNN. However, in the context of 3D medical images using 2D convolutions appear to be sub-optimal because valuable information along the third axis cannot be aggregated and taken into consideration, while, in videos, applying 2D convolutions on individual frames cannot well model the temporal information. On the other side, the computational cost for 3D CNN is large, making the deployment on edge devices difficult. To overcome the problem we decided to integrate some intuitions from the video action recognition field into the task of medical segmentation. We obtained a final network having computational complexity equal to 2D CNNs but performance comparable to fully 3D CNNs.

6. Results

We evaluated our model on AMOS [18] dataset introduced as part of the MICCAI 2022 challenge. AMOS is a

Models	mDSC(%)	Params(M)	Flops(G)
UNet [19]	88.87	31.18	680.31
VNet [20]	81.96	45.65	849.96
CoTr [21]	77.13	41.87	668.15
nnFormer [22]	85.63	150.14	425.78
UNETR [23]	78.33	93.02	177.51
Swin.UNETR [24]	86.37	62.83	668.15
Our Model	87.27	6.48	288.99

Table 1
Overall results of six state-of-the-art methods taken from the official AMOS-CT validation benchmark in [18] and our model.

large-scale, diverse, clinical dataset for abdominal organ segmentation that provides 500 CT and 100 MRI scans accompanied by voxel-level annotations for 15 organs. We compared our model with six state-of-the-art medical segmentation methods present in the benchmark in AMOS using the Dice Score as the evaluation metric. The Dice score is a common metric used to measure the amount of overlap between two regions. It ranges from 0 to 1, where 1 corresponds to a pixel-perfect match between the deep learning model output and ground truth annotation. As shown in Table 1 we achieve an overall accuracy of 87.27% gaining the second position in the benchmark right after UNet [19]. For a fair comparison, the results in the table are obtained by training for 1000 epochs using SGD optimizer with a momentum of 0.99, warm-up cosine scheduler for 50 iterations, an initial learning rate of 0.01, and a batch size of 2, recreating the same training condition of the benchmark created in [18]. We also expressed the model complexity in terms of floating-point operations per second (FLOPs) and the number of parameters. In the same table, we can see that our network is computationally and parameter count-wise more efficient by being on average ×2 faster with about 1/3 of parameters.

In Figure 1 we visualize representative samples comparing the groundtruth with our predictions. In the first row, we can see, as pointed out by the red arrow, that the segmentation masks for the pancreas (light green) and inferior vena cava (dark green) are separated but should touch one another. There are only a few pixels misclassified in this example, as in the last row where the segmentation mask of the pancreas highlighted in green is slightly larger than it should be. The larger error in the figure can be seen in the second row, where our model incorrectly labels parts of the duodenum highlighted in yellow, with the stomach, which is highlighted in blue.

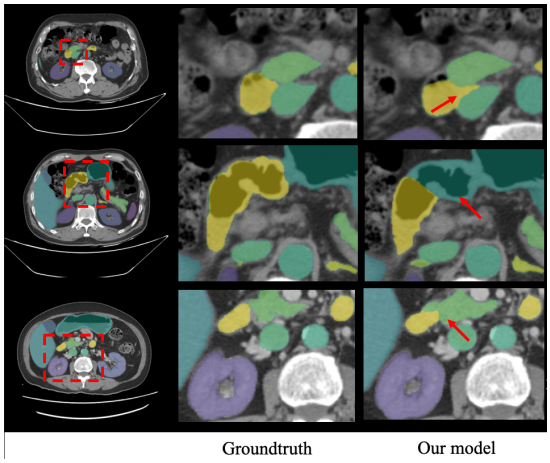

Figure 1: Qualitative visualizations of the proposed model on the AMOS-CT validation set.

7. Conclusion

We acknowledged the existence of a research gap in 3D anomaly detection also due to the lack of proper public datasets. To overcome the problem we decided to temporarily shift our attention to the segmentation of volumetric data. Our target was to develop an efficient model that can reach comparable results to other state-of-the-art approaches in the field, and we were able to obtain that. Our next target is to adapt the obtained model for defect detection using our internal industrial dataset.

References

[1] F. E. Grubbs, Procedures for detecting outlying observations in samples, Technometrics 11 (1969) 1–21.

[2] J. Yoon, K. Sohn, C.-L. Li, S. O. Arik, C.-Y. Lee, T. Pfister, Self-supervise, refine, repeat: Improving unsupervised anomaly detection, (2022).

[3] Z. Xiao, Q. Yan, Y. Amit, Do we really need to learn representations from in-domain data for outlier detection?, (2021).

[4] N. Shvetsova, B. Bakker, I. Fedulova, H. Schulz, D. V. Dylov, Anomaly detection in medical imaging with deep perceptual autoencoders, IEEE Access 9 (2021) 118571–118583.

[5] P. Perera, R. Nallapati, B. Xiang, Ocgan: One-class novelty detection using gans with constrained latent representations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 2898–2906.

[6] P. Bergmann, M. Fauser, D. Sattlegger, 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.

[7] P. Bergmann, X. Jin, D. Sattlegger, C. Steger, The mvtec 3d-ad dataset for unsupervised 3d anomaly detection and localization, (2021).

[8] E. Horwitz, Y. Hoshen, An empirical investigation of 3d anomaly detection and segmentation, (2022).

[9] J. Simarro Viana, E. de la Rosa, T. Vande Vyvere, D. Robben, D. M. Sima, C.-T. P. a. Investigators, Unsupervised 3d brain anomaly detection, in: A. Crimi, S. Bakas (Eds.), Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries, Springer International Publishing, 2021, pp. 133–142.

[10] M. Bengs, F. Behrendt, J. Krüger, R. Opfer, A. Schlaefer, Three-dimensional deep learning with spatial erasing for unsupervised anomaly segmentation in brain mri, International Journal of Computer Assisted Radiology and Surgery 16 (2021) 1413 – 1423.

[11] L. Wang, Y. Xiong, Z. Wang, Y. Qiao, D. Lin, X. Tang, L. Van Gool, Temporal segment networks: Towards good practices for deep action recognition, in: European conference on computer vision, Springer, 2016, pp. 20–36.

[12] C. Luo, A. L. Yuille, Grouped spatial-temporal aggregation for efficient action recognition, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 5512–5521.

[13] J. Lin, C. Gan, S. Han, Tsm: Temporal shift module for efficient video understanding, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 7083–7093.

[14] S. Sudhakaran, S. Escalera, O. Lanz, Gate-shift networks for video action recognition, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 1102–1111.

[15] U. Baid, S. Ghodasara, M. Bilello, S. Mohan, E. Calabrese, E. Colak, K. Farahani, J. Kalpathy-Cramer, F. C. Kitamura, S. Pati, L. M. Prevedello, J. D. Rudie, C. Sako, R. T. Shinohara, T. Bergquist, R. Chai, J. A. Eddy, J. Elliott, W. C. Reade, T. Schaffter, T. Yu, J. Zheng, B. Annotators, C. Davatzikos, J. T. Mongan, C. Hess, S. Cha, J. E. Villanueva-Meyer, J. B. Freymann, J. S. Kirby, B. Wiestler, P. Crivellaro, R. R.Colen, A. Kotrotsou, D. Marcus, M. Milchenko, A. Nazeri, H. M. Fathallah-Shaykh, R. Wiest, A. Jakab, M.-A. Weber, A. Mahajan, B. H. Menze, A. E. Flanders, S. Bakas, The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification, ArXiv abs/2107.02314 (2021).

[16] M. Antonelli, A. Reinke, S. Bakas, K. Farahani, A. Kopp-Schneider, B. A. Landman, G. Litjens, B. Menze, O. Ronneberger, R. M. Summers, et al., The medical segmentation decathlon, Nature communications 13 (2022) 1–13.

[17] B. Landman, Z. Xu, J. Igelsias, M. Styner, T. Langerak, A. Klein, Miccai multi-atlas labeling beyond the cranial vault–workshop and challenge, in: Proc. MICCAI Multi-Atlas Labeling Beyond Cranial Vault—Workshop Challenge, volume 5, 2015, p. 12.

[18] Y. Ji, H. Bai, J. Yang, C. Ge, Y. Zhu, R. Zhang, Z. Li, L. Zhang, W. Ma, X. Wan, et al., Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation, arXiv preprint arXiv:2206.08023 (2022).

[19] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer, 2015, pp. 234–241.

[20] F. Milletari, N. Navab, S.-A. Ahmadi, V-net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 fourth international conference on 3D vision (3DV), Ieee, 2016, pp. 565–571.

[21] Y. Xie, J. Zhang, C. Shen, Y. Xia, Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation, Medical Image Computing and Computer Assisted Intervention (2021) 171–180.

[22] H.-Y. Zhou, J. Guo, Y. Zhang, L. Yu, L. Wang, Y. Yu, nnformer: Interleaved transformer for volumetric segmentation, arXiv preprint arXiv:2109.03201 (2021).

[23] A. Hatamizadeh, Y. Tang, V. Nath, D. Yang, A. Myronenko, B. Landman, H. R. Roth, D. Xu, Unetr: Transformers for 3d medical image segmentation, in: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, 2022, pp. 574–584.

[24] Y. Tang, D. Yang, W. Li, H. R. Roth, B. Landman, D. Xu, V. Nath, A. Hatamizadeh, Self-supervised pre-training of swin transformers for 3d medical image analysis, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 20730–20740.