Hybrid Anomaly Detection for Industrial Images

Master research seminar at University of Zagreb, Faculty of Electrical Engineering and Computing

Dario Deković Matej Grcić Siniša Šegvić

***Abstract*—Dense anomaly detection in industrial images is an important task that can significantly reduce the costs and time of production. Most of the current methods rely on the information from neighboring pixels. We propose a method adaptation from the open set semantic segmentation task to a purely dense anomaly detection task. We present the adaptation in detail and describe the experiments. Finally, we describe the results and give directions for future research.**

***Index Terms*—dense anomaly detection, industrial imaging, hybrid anomaly detection,**

I. INTRODUCTION

Dense anomaly detection in industrial images is an important task in modern computer vision. It enables early fault detection and, consequently, reduces later costs related to quality assurance and product withdrawal. Many challenges arise with this task, which is mainly related to a variety of objects in images and possible anomalies. One of the main challenges is the lack of anomaly labels in the training dataset. In this work, we propose an adaptation of a novel method for dense open set recognition DenseHybrid [1] for a pure dense anomaly detection task. We decide to simply name it DenseHybridAD. This method combines both generative and discriminative approaches to anomaly detection thus minimizing their pitfalls. The main challenges of this adaptation lie in the lack of a primary discriminative task in most anomaly detection datasets. To address this we propose using a pretrained network to mitigate the information loss. In the first part of this paper, we give a short overview of the current state of the art in dense anomaly detection for industrial imaging. After, we give a detailed description of our approach and present the results on a well know MVTec AD [2] dataset.

II. RELATED WORK

Most of the modern dense anomaly detection methods fall into either a discriminative or generative category. Furthermore, in industrial imaging, most of the top methods use information from neighboring pixels to detect anomalies. State-of-the-art method on MVTec AD dataset [3] uses a histogram of representative features to make use of positional information. Another method described in [4] uses neighboring pixel features to model the nominal distribution of the target pixel. Our method differs from this in that it doesn't look at the specific neighborhood of a certain pixel but rather tries to model the underlying data distribution and blend it with a discriminative dense anomaly prediction.

III. METHOD AND IMPLEMENTATION

Our approach builds on the hybrid approach described in the paper [1]. Thus our loss function has both a generative and a discriminative component. This allows us to exploit the benefits from both approaches. A general overview of the method is described in figure 1 and more details about each component will be given in the following section. The notation in figure 1 will be used throughout the paper.

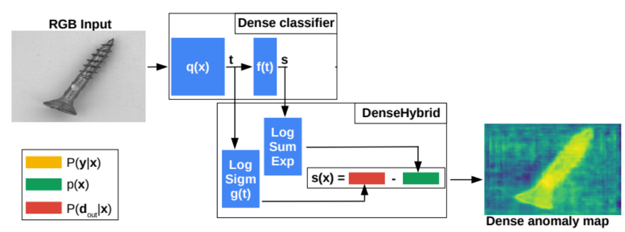

Figure 1. Visualization of the DenseHybridAD method. Part of this image was taken from [1].

A. Hybrid anomaly detection

The first part of our loss function is a loss $L_T H$ from the second head g of our model which is tasked with modeling a probability that a given pixel x is an outlier. The second part is a reinterpretation of the logits from a segmentation model as an unnormalized joint log-distribution of classes and data. This we model as an energy loss L_x. Finally, to address the lack of anomalies in the training set, we simulate the out-of-distribution pixels D_{out} with negative images that we paste onto our training images. By combining these two losses we arrive at a modified DenseHybrid [1] loss adjusted for the lack of a primary discriminative task.

$$\begin{aligned} L_{AD} = &- \mathbb{E}_{x,y \in D_{in}}[\ln P(d_{in}|x)] \\ &- \beta \cdot \mathbb{E}_{x \in D_{out}}[\ln P(d_{out}|x) - \ln p(x)] \end{aligned} \tag{1}$$

However, we can assume that the lack of the primary discriminative loss will heavily influence L_x part of our L_{AD}. This was also empirically shown in the experiments we conducted. For this, we propose two main solutions which are crucial for the adaptation of the DenseHybrid method to the problem of dense anomaly detection.

Firstly, we postulate that it is necessary to have a model pretrained on a semantic segmentation task. For our purposes, we choose a DeepLabV3 model [5] with a ResNet-50 backbone [6]. We initialize the model with weights pretrained on the PascalVOC dataset [7]. We freeze some of the layers of the model which will be described in more detail in the experiment section of the paper. With this we aim to achieve a smaller influence of the energy loss on our model thus, keeping its good properties related to the modeling of the underlying data distribution.

Secondly, we propose that negative images are not pasted as a whole but rather that the individual class masks are taken out and pasted as anomalies. With this, we try to prepare our model for the wide variety of shapes of possible anomalies found in industrial images. Furthermore, this is another way to mitigate the effect of the lack of discriminative loss on our energy loss. In the original DenseHybrid paper rectangular negative patches were sufficient since the model had the shape information from the discriminative task.

IV. Experiments

We conduct experiments on the MVTec AD dataset [2] using the above described methods. We validate our assumptions about the effect of freezing some layers of a pretrained model and using mask-shaped patches. For the negative images we choose the ADE20K dataset [8]. All of the experiments were conducted over 40 epochs with the β hyperparameter of the loss set at 0.03. When we describe freezing the layers *All* denotes freezing all the layers except the final BN-ReLU-Conv layer of both heads.

Table I
Anomaly detection on MVTec AD dataset [2] using DenseHybridAD method.

Freeze layers	Patch type	AUROC [%]	FPR_{95}[%]	AP [%]
All	Mask	80.72	28.25	11.49
Backbone	Mask	73.31	43.1	11.59
None	Rectangular	72.53	38.46	7.71
All	Rectangular	74.03	34.93	8.51

From the table I we can see how our assumptions had a positive impact on the performance. Also, the freezing of the layers had the bigger impact from the two approaches. In the following images, we will show an example in which our best model performed very well and one example in which it performed poorly.

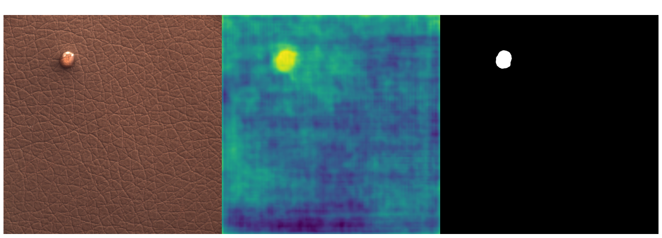

Figure 2. Example of well diagnosed anomaly. AUROC: 99.98%

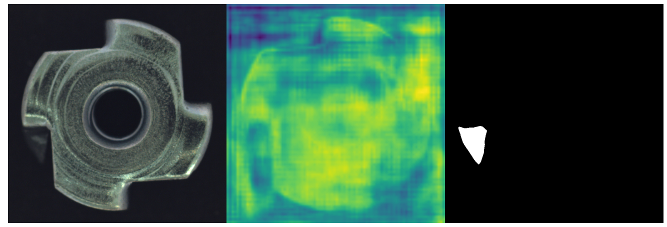

Figure 3. Example of a poorly diagnosed anomaly. AUROC: 14.21%

These images are exemplary of the trend we noticed in our experiments. When an anomaly looks like an individual class on a primary object our model performs very well. However, when an anomaly is some distortion of the primary object and doesn't have clear outlines of a separate class our model doesn't perform well. This can be explained by the fact that we are using a pretrained model and freezing a majority of its layers thus preventing the model from deviating from its primary discriminative task too much. We show that this is very beneficial to the overall performance, on the other hand, because of this our model fails to capture the nuances of the new dataset.

Finally, in table II we present comparative results of our method to the state of the art on MVTec AD.

Table II
Comparative results of our method with the SOTA on MVTec AD.

Method	AUROC [%]	Extra training data
DenseHybridAD (ours)	80.72	✓
PNI Ensemble [3]	99.05	✓
MemSeg [9]	98.84	✗
MemSeg [9]	98.75	✗

V. Conclusion

In this paper, we proposed an adapted method from the dense open-set recognition world to pure dense anomaly detection. We outlined the necessary changes to the method to achieve better performance. We have shown in our experiments how our changes are valid and that they indeed result in better performance. However, even these results are still very far behind the state of the art. This can be explained by the fact that our method was intended to work with the primary discriminative task and our changes are not enough to mitigate the resulting loss of information. On the other hand, we believe our paper warrants further research into the topic. A future research direction could be mixing the batches from the MVTec AD dataset with some well-known dataset to make use of its discriminative task.

Literatura

[1] M. Grcic, P. Bevandi'c, and S. vSegvi'c, "Densehybrid: Hybrid anomaly detection for dense open-set recognition," in *European Conference on Computer Vision*, 2022.

[2] P. Bergmann, M. Fauser, D. Sattlegger, and C. Steger, "Mvtec ad — a comprehensive real-world dataset for unsupervised anomaly detection," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9584–9592, 2019.

[3] J. Bae, J. Lee, and S. Kim, "Image anomaly detection and localization with position and neighborhood information," *ArXiv*, vol. abs/2211.12634, 2022.

[4] J. Jang, E. Hwang, and S.-H. Park, "N-pad : Neighboring pixel-based industrial anomaly detection," *ArXiv*, vol. abs/2210.08768, 2022.

[5] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *ArXiv*, vol. abs/1706.05587, 2017.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2015.

[7] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The pascal visual object classes (voc) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303–338, Jun. 2010.

[8] B. Zhou, H. Zhao, X. Puig, S. Fidler, A. Barriuso, and A. Torralba, "Semantic understanding of scenes through the ade20k dataset," *International Journal of Computer Vision*, vol. 127, pp. 302–321, 2016.

[9] M. Yang, P. Wu, J. Liu, and H. Feng, "Memseg: A semi-supervised method for image surface defect detection using differences and commonalities," *ArXiv*, vol. abs/2205.00908, 2022.