Machine Learning and Computer Vision for Anomaly Detection in Industrial Settings

Axel De Nardin, Pankaj Mishra, Claudio Piciarelli, Gian Luca Foresti
Dipartimento di Scienze Matematiche, Informatiche e Fisiche, Università degli Studi di Udine, Via delle Scienze, 206, 33100 Udine, Italia
{denardin.axel,mishra.pankaj}@spes.uniud.it, {claudio.piciarelli,gianluca.foresti}@uniud.it,

Abstract

The problem of image anomaly detection is of uttermost importance for many applications, one of these being represented by quality control and anomaly detection in Industrial production lines. This task is typically performed manually by human operators who inspect the products directly or through an online video capture system, but recently the adoption of automated systems based on Deep Learning models is becoming more and more widespread. In this work, we propose various approaches which aim to solve the problems of Anomaly Detection and Localization in such away.

1 Introduction

The ability to identify anomalous instances in large sets of data is of great importance in many different fields of application. One typical example is represented by the need to detect and discard faulty products in industrial production lines. The ability to do so in an automated and effective way represents a very important problem for manufacturing companies as it can make the difference between gaining a profit and suffering a loss. Furthermore, the inability to detect faulty components in some contexts may lead to producing potentially dangerous products.

While for humans dealing with this kind of problem is a rather easy task, we cannot say the same for machines. One main reason that makes it a difficult problem to address is the fact that, while it is essentially a classification problem, classical classification approaches cannot be used because of the nature of the data analyzed. In fact, when dealing with anomaly detection problems, the data is often highly unbalanced in favor of "normal" instances, while we have very few examples representing the "abnormal" ones. When dealing with images we also have the added problem of the high dimensionality of the data which often leads to more classical methods for anomaly detection, such as clustering techniques, to achieve poor performance. Finally, another very common problem that makes it hard to detect and localize anomalies in images, is represented by the heterogeneity in their appearance and in the fact that sometimes they can be very subtle, making them particularly difficult to identify.

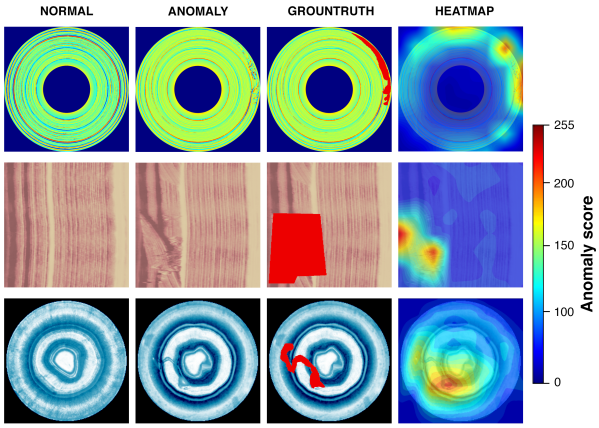

Figura 1: BTAD dataset. First Col: normal images; Second col: anomalous images; Third col: anomalous images with ground truth; Fourth col: heat map and anomaly score generated by our method VTADL.

For this reason, the aim of the present work is to propose a variety of robust deeplearning-based approaches which are able to effectively tackle the problems of detection and localization of anomalies in images. Our main focus regards particular reconstruction-based methods, in which the original image, given input, is processed by a deep learning model which first tries to extract its most representative features and then leverages them in order to reconstruct it as accurately as possible. Finally, the two images are compared and a distance measure between them is calculated. The idea behind this family of approaches is that, if the model is trained only on "normal" images when an anomalous one is provided to it, its reconstruction will be less accurate and therefore the value of the distance measure will be larger compared to the value obtained for images without anomalies, making it possible to discriminate between the two.

2 Anomaly detection and localization

The process adopted in order to perform anomaly detection and localization on images, as briefly introduced in the previous sections, relies on the comparison between the original image and one reconstructed from it. Typically a distance

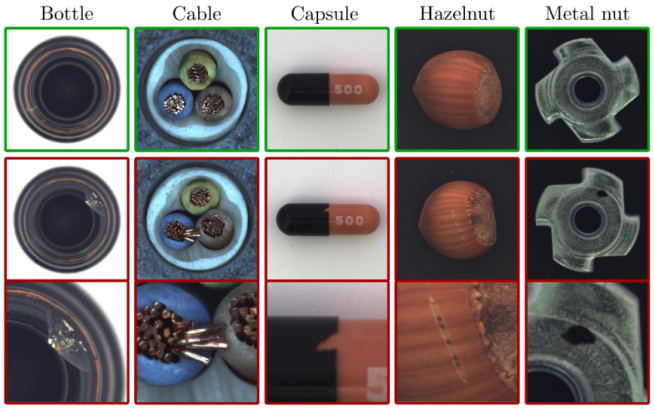

Figura 2: Samples for 5 of the classes present in the MVTec dataset. For each of them is provided a normal instance (top row), an instance containing an anomaly (middle row), and a zoomed view on the defect (bottom row)

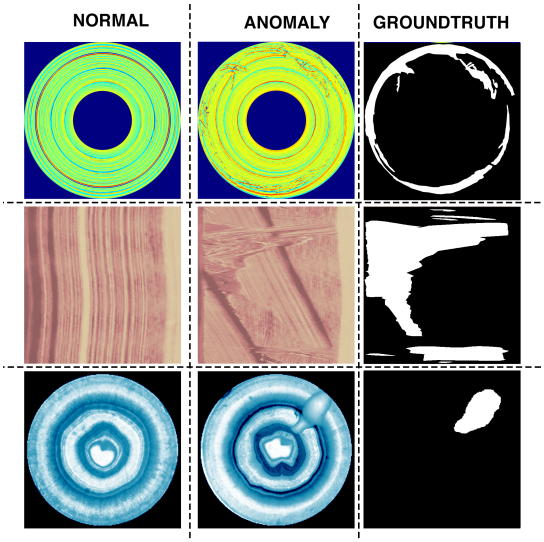

Figura 3: BTAD dataset. First column: normal images pf thee industrial products; Second column: anomalous images; Third column: pixel-precise ground truth

metric is calculated between each pixel composing the images, representing how different the two are. Then a threshold value can be introduced to discriminate between normal and anomalous regions of each instance of the dataset. This process is referred to as "Anomaly Localization". On the other hand the process of "Anomaly Detection" consists in identifying anomalies at an instance level or, in other words, identifying those instances that contain an anomalous region. To do so a typical approach is to average out the distance measure between each of the pixels composing an image in order to obtain a single value that describes the anomaly score of the image itself.

2.1 Benchmark

A popular dataset which in recent years has become the defacto standard as a benchmark for anomaly detection and localization approaches is represented by MVTec [Bergmann *et al.*, 2020]. This dataset consists of 3629 training images and 1725 testing images divided into 15 classes (a sample is provided in Fig. 2), five of which represent different textures and the remaining ones covering a set of products with heterogeneous characteristics, some of them present a rigid structure, others are deformable or present natural variations in their appearance. Furthermore, the way in which the images are captured is also heterogeneous, for some of the classes all the instances belonging to the present the product in a roughly aligned fashion while for some others a random rotation is introduced. In total 73 different types of defects are provided, 5 for each category on average. To make the testing procedure possible pixel-accurate labels for all the defective regions in each image are also provided by the authors.

Another recent benchmark dataset BTAD (the beanTech Anomaly Detection dataset) has been published by [Mishra *et al.*, 2021]. The dataset contains a total of 2830 real-world RBG images of 3 industrial products showcasing body and surface defects. It also contains pixel-precise ground truth of the three industrial products (see Fig 3). Product 1 is 1600×1600 pixels, product 2 is 600×600 and product 3 is 800×600 pixels in size. Product 1, 2, and 3 have 400, 1000, and 399 train images respectively.

3 Deep Learning Approaches

In recent years Deep Learning approaches took over more classical Computer Vision techniques in the context of anomaly detection and localization. One of the main reasons behind this is that they provide an endtoend framework capable of automatizing the process of feature selection, previously done explicitly, and at the same time solving the main problem for which they are designed. The most popular models adopted in this area are represented by Deep Neural Networks employing Convolutional Layers, mainly AutoEncoders (AE) [Bergmann *et al.*, 2020], and Generative Adversarial Networks (GANs) [Schlegl *et al.*, 2019] based ones. The recent works, presented in the present work, on the other hand, focus on a newly proposed architecture called Vision Transformer (ViT) [Dosovitskiy *et al.*, 2021] which introduces the concept of selfattention (a concept typically applied in NLP applications) between different regions of the processed images. The main idea is that each patch of the input image gets mapped by the network into an embedding which is highly dependant on the other regions composing a said image, allowing therefore to produce a description of each of the patches which includes a large amount of contextual information, instead of considering only its direct surroundings as typically happens when using convolution based networks.

Talking in terms of learning most of the deep learning methods are categorized into either

- Supervised learning: An deep learning approach where we have training labels along with the training dataset.
- semi-supervised learning: An deep learning approach where we have either limited labels or class-imbalanced labels for the training dataset.

- Unsupervised learning: An deep learning approach where we don't have training labels available with the training dataset.

In our research work, where we deal with anomaly detection in images and try to solve real-life industrial quality inspection tasks. We developed methods in all three above-mentioned deep learning approaches. A brief discussion of them can be seen below sections.

3.1 Supervised Global Anomaly Classification

Our supervised work [Piciarelli *et al.*, 2019] with adapted capsule network [Sabour *et al.*, 2017], solved the problem of high-imbalance data training in real-life cases. Our proposed method currently outperforms or is comparable to other state-of-the-art methods, however, a direct comparison would be unfair since most of those methods use semi-supervised or unsupervised techniques. We proposed an alternative approach based on fully supervised learning with imbalanced datasets. This idea came from real-world scenarios, in which anomalous data are often available but their amount is extremely scarce. The proposed approach, which is a variant of the capsnet [Sabour *et al.*, 2017] architecture, showed good performances even with extremely imbalanced datasets, outperforming both the standard capsnet architecture and other anomaly detection techniques.

3.2 Semi-supervised Approaches for Global Image Classification

In addition to the supervised approach [Piciarelli *et al.*, 2019] [Piciarelli *et al.*, 2021], we also developed novel semi-supervised approaches for the global image anomaly classification. We proposed deep pyramidal network with stacked autoencoders for anomaly detection [Mishra *et al.*, 2020a]. Anomalies are identified by means of a network that encodes normal images in a low-dimensional latent space and then reconstructs them, ideally modeling an identity function. Since the network is trained on normal data only, it fails at reconstructing anomalous images, which can be detected by an image similarity loss. The main contributions of this work consist in the usage of a multi-scale pyramidal approach that extracts latent features at different resolutions, and the usage of a high-level perceptual loss to better compare images at the feature level, rather than at pixel level. Achieved results are promising and often outperform other state-of-the-art methods.

Moreover, we proposed another novel network Pyramidal Image Anomaly Detector (PIADE) [Mishra *et al.*, 2020b], a deep reconstruction-based pyramidal approach, in which image features are extracted at different scale levels to better catch the peculiarities that could help to discriminate between normal and anomalous data. The network is trained on normal data only, and it builds a "normality model" by mapping the input images in a low-dimension feature space, from which they can be correctly reconstructed. The inability of the network to reconstruct anomalous images allows the identification of anomalies, which can be detected by their higher reconstruction error. Compared to other state-of-the-art works, the proposed models include a pyramidal multi-scale approach to analyze image features at different scale levels, a dynamic routing layer inspired by the architecture of capsule networks[Sabour *et al.*, 2017], and a high-level image comparison loss. Moreover, the system has been tested not only on standard datasets such as CIFAR10 and COIL-100 [Nene *et al.*, 1996] (which have not been initially created for anomaly detection experiments) but also on the recently proposed MVTec dataset of anomalies in industrial images. Experimental results showed that the proposed model is at-par, and often outperforms other state-of-the-art works.

3.3 Unsupervised Approaches for Global Image Classification and Localization

We proposed a transformer-based framework a Vision Transformer Network for Image Anomaly Detection and Localization (VT-ADL) [Mishra *et al.*, 2021], which uses reconstruction and patch-based learning for image anomaly detection and localization. The anomalies can be detected at a global level using a reconstruction-based approach and can be localized with the application of a Gaussian mixture model applied to the encoded image patches. The achieved results are at par with or outperform other state-of-the-art techniques. We also published BTAD3, a real-world industrial dataset for the anomaly detection task.

4 Experimental Results

In this section, we present some of the results obtained through our proposed novel methods. Figure 1 and 4 shows the heat-map and anomaly score generated by our method VT-ADL. This approach doesn't need any ground truth for training and produces pixel-wise anomaly score for anomaly localization and an overall score for anomaly classification of the images.

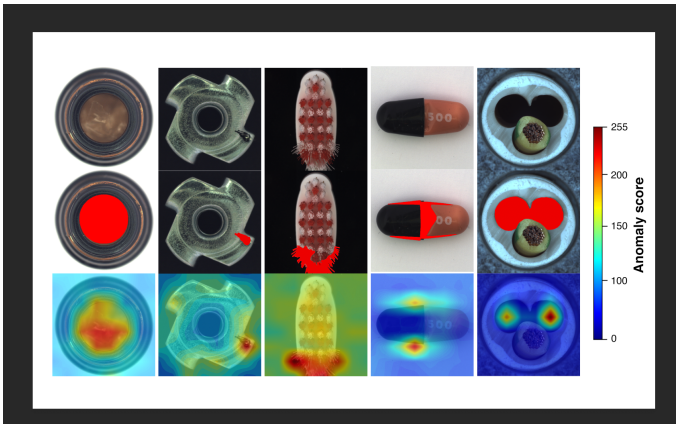

Figura 4: Anomaly detection on MVTec dataset. The first row: the anomalous image of the bottle, cable, capsule, metal nut, and brush; Second row: ground truth; Third row: generated anomaly score and anomaly localization by our method VT-ADL

Some other comparative studies with other state-of-the-arts like FCN32 [Shelhamer *et al.*, 2017], Unet [Ronneberger *et al.*, 2015], Unet++ [Zhou *et al.*, 2019], UNet2 [Jiao *et al.*, 2020], SegNet [Badrinarayanan *et al.*, 2017], in terms of

numbers of parameter and inference time can be seen in the figure

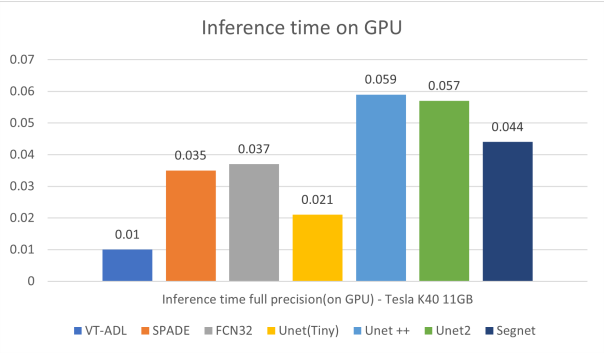

Figura 5: Inference time (in sec) of full precision models over GPU from all the deep models.

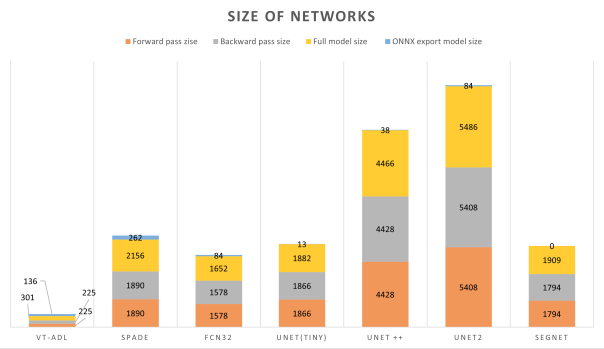

Figura 6: Network Sizes: The stacked bar plot shows the size (in MB) of the forward pass, backward pass, full-size model, and the ONNX-exported model size of the deep models used in this study.

Riferimenti bibliografici

[Badrinarayanan *et al.*, 2017] Vijay Badrinarayanan, Alex Kendall, e Roberto Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE transactions on pattern analysis and machine intelligence*, 39(12):2481–2495, 2017.

[Bergmann *et al.*, 2020] Paul Bergmann, Michael Fauser, David Sattlegger, e Carsten Steger. Uninformed students: Student-teacher anomaly detection with discriminative latent embeddings. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4183–4192, 2020.

[Dosovitskiy *et al.*, 2021] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, e Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021.

[Jiao *et al.*, 2020] Libin Jiao, Lianzhi Huo, Changmiao Hu, e Ping Tang. Refined unet v2: End-to-end patch-wise network for noise-free cloud and shadow segmentation. *Remote Sensing*, 12(21):3530, 2020.

[Mishra *et al.*, 2020a] Pankaj Mishra, Claudio Piciarelli, e Gian Luca Foresti. Image anomaly detection by aggregating deep pyramidal representations, 2020.

[Mishra *et al.*, 2020b] Pankaj Mishra, Claudio Piciarelli, e Gian Luca Foresti. A neural network for image anomaly detection with deep pyramidal representations and dynamic routing. *International Journal of Neural Systems*, 30(10):2050060–2050060, 2020.

[Mishra *et al.*, 2021] Pankaj Mishra, Riccardo Verk, Daniele Fornasier, Claudio Piciarelli, e Gian Luca Foresti. VT-ADL: A vision transformer network for image anomaly detection and localization. In *30th IEEE/IES International Symposium on Industrial Electronics (ISIE)*, June 2021.

[Nene *et al.*, 1996] Sameer A. Nene, Shree K. Nayar, e Hiroshi Murase. object image library (coil-100. Technical report, 1996.

[Piciarelli *et al.*, 2019] Claudio Piciarelli, Pankaj Mishra, e Gian Luca Foresti. Image anomaly detection with capsule networks and imbalanced datasets. In *International Conference on Image Analysis and Processing*, pages 257–267. Springer, 2019.

[Piciarelli *et al.*, 2021] Claudio Piciarelli, Pankaj Mishra, e Gian Luca Foresti. Supervised anomaly detection with highly imbalanced datasets using capsule networks. *International Journal of Pattern Recognition and Artificial Intelligence*, page 2152010, 2021.

[Ronneberger *et al.*, 2015] Olaf Ronneberger, Philipp Fischer, e Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[Sabour *et al.*, 2017] Sara Sabour, Nicholas Frosst, e Geoffrey E Hinton. Dynamic routing between capsules. In *Advances in neural information processing systems*, pages 3856–3866, 2017.

[Schlegl *et al.*, 2019] Thomas Schlegl, Philipp Seeböck, Sebastian Waldstein, Georg Langs, e Ursula Schmidt-Erfurth. f-anogan: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54, 01 2019.

[Shelhamer *et al.*, 2017] Evan Shelhamer, Jonathan Long, e Trevor Darrell. Fully convolutional networks for semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4):640–651, 2017.

[Zhou *et al.*, 2019] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, e Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.