Semi-supervised Anomaly Detection using AutoEncoders

Manpreet Singh Minhas
University of Waterloo, ON, Canada
John Zelek
University of Waterloo, ON, Canada
Email: {msminhas,jzelek}@uwaterloo.ca

Abstract

Anomaly detection refers to the task of finding unusual instances that stand out from the normal data. In several applications, these outliers or anomalous instances are of greater interest compared to the normal ones. Specifically in the case of industrial optical inspection and infrastructure asset management, finding these defects (anomalous regions) is of extreme importance. Traditionally and even today this process has been carried out manually. Humans rely on the saliency of the defects in comparison to the normal texture to detect the defects. However, manual inspection is slow, tedious, subjective and susceptible to human biases. Therefore, the automation of defect detection is desirable. But for defect detection lack of availability of a large number of anomalous instances and labelled data is a problem. In this paper, we present a convolutional auto-encoder architecture for anomaly detection that is trained only on the defect-free (normal) instances. For the test images, residual masks that are obtained by subtracting the original image from the auto-encoder output are thresholded to obtain the defect segmentation masks. The approach was tested on two data-sets and achieved an impressive average F1 score of 0.885. The network learnt to detect the actual shape of the defects even though no defected images were used during the training.

1 Introduction

An anomaly is anything that deviates from the norm. Anomaly detection refers to the task of finding the anomalous instances. Defect detection is a special case of anomaly detection and has applications in industrial settings. Manual inspection by humans is still the norm in most of the industries. The inspection process is completely dependent on the visual difference of the anomaly (defect) from the normal background or texture. The process is prone to errors and has several drawbacks, such as training time and cost, human bias and subjectivity, among others. Individual factors such as age, visual acuity, scanning strategy, experience, and training impact the errors caused during the manual inspection process [1]. As a result of these challenges faced in the manual inspection by humans, automation of defect detection has been a topic of research across different application areas such as steel surfaces [2], rail tracks [3] and fabric [4], to name a few. However, all these techniques face two common problems: lack of large labelled data and the limited number of anomalous samples. Semi-supervised techniques try to tackle this challenge. These techniques are based on the assumption that we have access to the labels for only one class type i.e. the normal class [5]. They try to estimate the underlying distribution of the normal samples either implicitly or explicitly. This is followed by the measurement of deviation or divergence of the test samples from this distribution to determine an anomalous sample. To take an example of semi-supervised anomaly detection, Schlegl et al. [6] used Generative Adversarial Networks (GANs) for anomaly detection in optical coherence tomography images of the retina. They trained a GAN on the normal data to learn the underlying distribution of the anatomical variability. But they did not train an encoder for mapping the input image to the latent space. Because of this, the method needed an optimization step for every test image to find a point in the latent space that corresponded to the most visually similar generated image which made it slow. In this research, we explore an auto-encoder based approach that also tries to estimate the distribution of the normal data and then uses residual maps to find the defects. It is described in the next section.

2 Method

The proposed network architecture is shown in Figure 1. It is similar to the UNet [7] architecture. The encoder (layers x1 to x5) uses progressively decreasing filter sizes from 11×11 to 3×3. This decreasing filter size is chosen to allow for a larger field of view for the network without having to use large number of smaller size filters. Since deeper networks have a greater tendency to over-fit to the data and have poor generalization. The decoder structure has kernel sizes that are in the reverse of the encoder order and uses Transposed Convolution Layers. The output from the encoder layers is concatenated with the previous layers before passing to layers x7 to x9. For every Conv2D(Transpose) layer the parameters shown are kernel size, stride and number of filters for that layer. After every layer, batch normalization [8] is applied which is followed by the ReLU activation function [9]. For a $H \times W$ input the network outputs a $H \times W$ reconstruction. The network is trained on only the defect-free or normal data samples. Tensorflow 2.0 was used for conducting the experiments. The loss function used was the L2 norm or MSE (Mean Squared Error). The label in this case is the original input image and the prediction is the image reconstructed by the auto-encoder. Adam optimizer [10] was used with default settings. The training was done for 50 epochs.

Our hypothesis is that the auto-encoder will learn representations that would only be able to encode and decode the normal samples properly and will not be able to reconstruct the anomalous regions. This shall cause large residuals for the defective regions in the residual map obtained by subtracting the reconstructed image from the input image as shown in Equation 1. The subtraction is done at per pixel-level. This is followed by a thresholding operation to obtain the final defect segmentation.

$$R = X - AE(X) \tag{1}$$

where R is the residual, X is the input and $AE(X)$ is the output (reconstructed image) of the auto-encoder. The data-sets used for conducting the experiments are described next.

3 Data-sets

1. **DAGM**[11] is a synthetic data-set for industrial optical inspection and contains ten classes of artificially generated textures with anomalies. For this study, the Class 8 having the crack defect was randomly selected. It (hereafter referred to as DAGMC8) contains 150 images with one defect per image and 1000 defect-free images.

2. **RSDDs (Rail surface discrete defects)** [12] contains varying sized images of two different types of rails. We randomly selected the RSDDs Type-I category (referred to as RSDDsI) containing 67 images from express rails for the experiments. Segmentation masks were available which were used to extract 200×160 patches from the images and were classified into the anomaly and normal class to build the training and test data-set.

4 Results

A few examples of segmentation results obtained after applying thresholding operation to the residual maps on the DAGMC8 and RSDDsI data-set are shown in 2 (a) and (b) respectively and the F1 score values are shown in Table 1. As can be seen, on the synthetic DAGMC8 data-set the network could detect the cracks (anomalies) and there was little to no noise. This is in concurrence with the high F1 score value of 0.96. However, for the RSDDsI data-set, the results are a bit noisy. The same point is reflected by a lower F1 Score value of 0.81. Even though the thresholded residual maps managed to detect the defects in most of the images, the results contained more noise. One more observation was that the segmentation result was very sensitive to the choice of the threshold and minor changes led to large variations in the detection output. For the RSDDsI data-set, the illumination conditions were also varying in addition to the inherent noise in the data-set. For some images in the data-set, these areas were also not properly reconstructed by

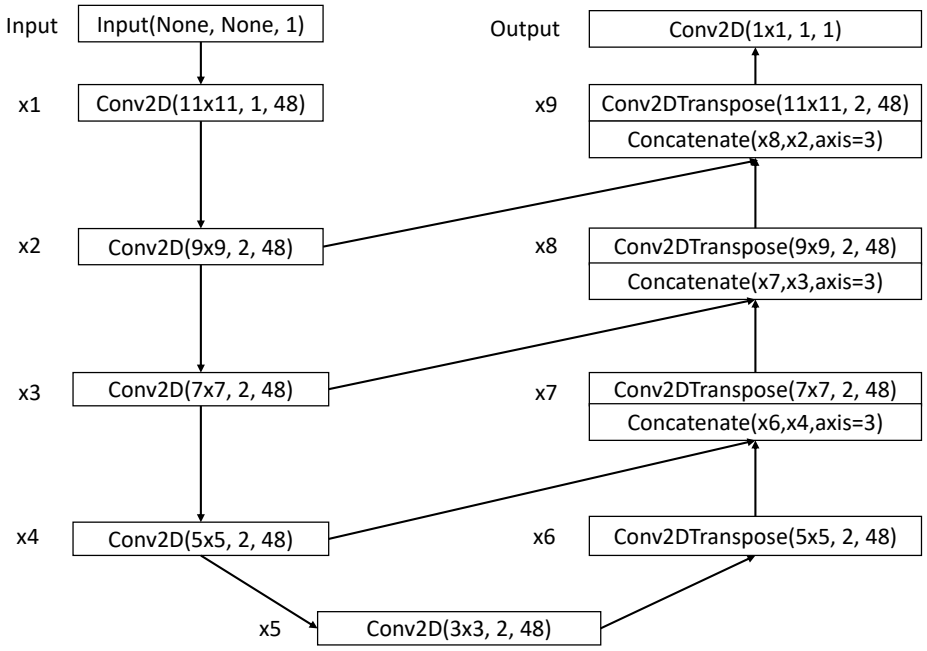
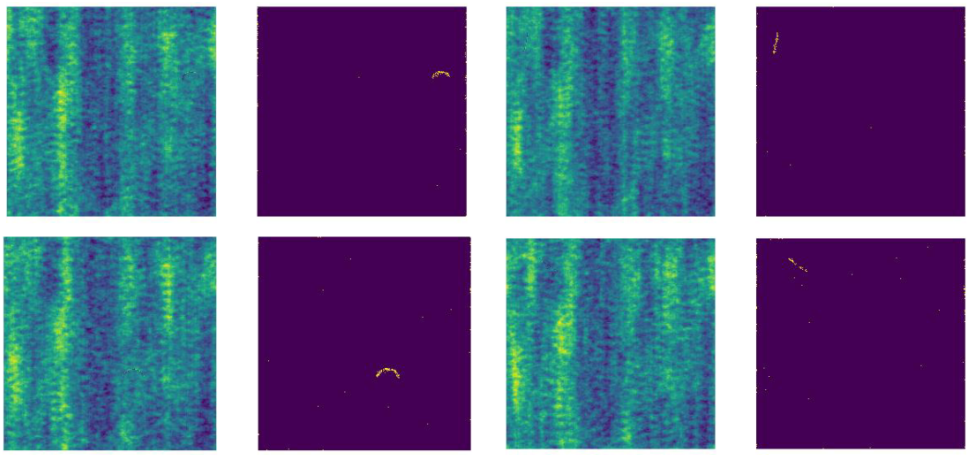
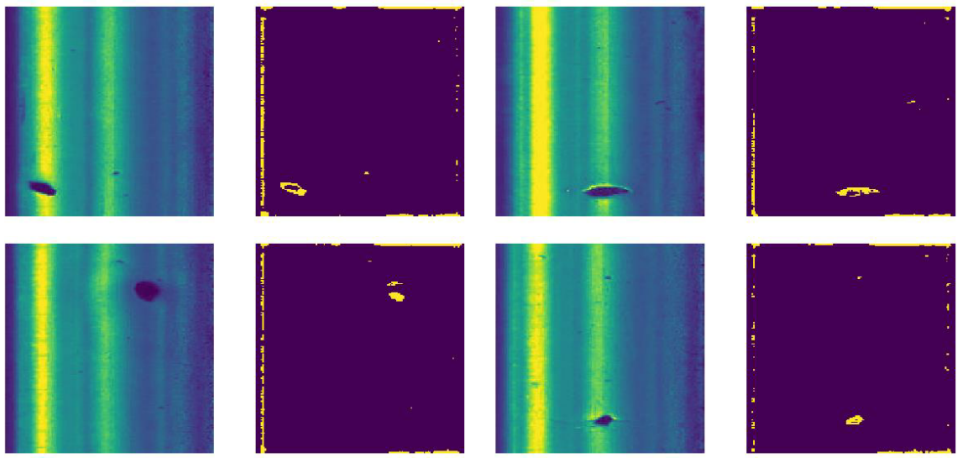


Fig. 1: Proposed auto-encoder network architecture (similar to the UNet architecture). The encoder (layers x1 to x5) uses progressively decreasing filter sizes from 11×11 to 3×3. The decoder structure has kernel sizes that are in the reverse of the encoder order and uses Transposed Convolution Layers. The output from the encoder layers is concatenated with the previous layers before passing to layers x7 to x9. For every Conv2D(Transpose) layer the parameters shown are kernel size, stride and number of filters for that layer. After every layer, batch normalization is applied which is followed by the ReLU activation function. For a $H \times W$ input the network outputs a $H \times W$ reconstruction. The network is trained on only the defect-free or normal data samples.



(a)



(b)

Fig. 2: Few examples of defect detection output on the DAGMC8 (Fig. 2 (a)) and RSDDsI (Fig. 2 (b)) data-sets respectively.

the auto-encoder, leading to false positives. Even though the geometric shapes and extent of the defects were different across images, the actual shapes of the anomalies were detected. This could be beneficial for applications where certain specific metrics need to be calculated for the defects. However, the lack of control over the types of defects that are detected by the auto-encoder reduces the targeting capability in comparison to supervised approaches.

Table 1: F1 Score values on the DAGMC1 and RSDDsI data-set obtained by using the proposed method.

Data-set	F1 Score
DAGMC8	0.96
RSDDsI	0.81

5 Conclusion and Future Work

In this work, we explored and presented a semi-supervised anomaly detection technique using deep learning based AutoEncoders. The proposed network architecture is similar to UNet. It can be trained using only the normal samples. This is an important feature that is essential for practical applications where a limited number of anomalous samples and a large number of normal samples are available. The approach led to an impressive average F1 score of 0.885 on two data-sets. Qualitative results obtained on two data-sets show that the technique leads to the detection of anomalies which can vary in terms of shape, geometry, etc. However, the method is sensitive to the choice of the threshold. Even illumination changes were picked up by the method as anomalies which is undesirable. For future work, experiments on data-sets with more than one defect type per image could be conducted. Structural Similarity Index (SSIM) could be explored as a loss function. It compares two images based on luminance, contrast, and structure and as a result is a better measure of visual similarity in comparison to the mean squared error. Also, some kind of statistic such as the L1 norm calculated on the residual images could be used as an anomaly score. Rather than subtracting the reconstructed image, other comparison methods should be explored. Exploring ways to make the network invariant to irrelevant factors such as illuminance needs to also be explored.

Acknowledgments

We thank the Ontario Ministry of Transportation and NSERC (National Science and Research Council) for providing funds that supported this research.

References

[1] J. E. See, C. G. Drury, A. Speed, A. Williams, and N. Khalandi, "The role of visual inspection in the 21st century," *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 61, no. 1, pp. 262–266, 2017. [Online]. Available: https://doi.org/10.1177/1541931213601548

[2] X. Sun, J. Gu, S. Tang, and J. Li, "Research progress of visual inspection technology of steel products — a review," *Applied Sciences*, vol. 8, no. 11, 2018. [Online]. Available: http://www.mdpi.com/2076-3417/8/11/2195

[3] H. Yu, Q. Li, Y. Tan, J. Gan, J. Wang, Y. Geng, and L. Jia, "A coarse-to-fine model for rail surface defect detection," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 3, pp. 656–666, March 2019.

[4] A. Kumar, "Computer-vision-based fabric defect detection: A survey," *IEEE Transactions on Industrial Electronics*, vol. 55, no. 1, pp. 348–363, Jan 2008.

[5] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Computing Surveys (CSUR)*, vol. 41, no. 3, p. 15, 2009. [Online]. Available: http://scholar.google.de/scholar.bib?q=info:jAfBmk-9uAcJ:scholar.google.com/&output=citation&hl=de&ct=citation&cd=0

[6] T. Schlegl, P. Seeböck, S. M. Waldstein, U. Schmidt-Erfurth, and G. Langs, "Unsupervised anomaly detection with generative adversarial networks to guide marker discovery," in *Information Processing in Medical Imaging*, M. Niethammer, M. Styner, S. Aylward, H. Zhu, I. Oguz, P.-T. Yap, and D. Shen, Eds. Cham: Springer International Publishing, 2017, pp. 146–157.

[7] O. Ronneberger, P.Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, ser. LNCS, vol. 9351. Springer, 2015, pp. 234–241, (available on arXiv:1505.04597 [cs.CV]). [Online]. Available: http://lmb.informatik.uni-freiburg.de/Publications/2015/RFB15a

[8] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proceedings of the 32Nd International Conference on International Conference on Machine Learning - Volume 37*, ser. ICML'15. JMLR.org, 2015, pp. 448–456. [Online]. Available: http://dl.acm.org/citation.cfm?id=3045118.3045167

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017. [Online]. Available: http://doi.acm.org/10.1145/3065386

[10] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, cite arxiv:1412.6980Comment: Published as a conference paper at the 3rd International Conference for Learning Representations, San Diego, 2015. [Online]. Available: http://arxiv.org/abs/1412.6980

[11] T. H. Matthias Wieler, "Weakly supervised learning for industrial optical inspection," https://hci.iwr.uni-heidelberg.de/node/3616.

[12] J. Gan, Q. Li, J. Wang, and H. Yu, "A hierarchical extractor-based visual rail surface inspection system," *IEEE Sensors Journal*, vol. 17, no. 23, pp. 7935–7944, Dec 2017.