

Improving Unsupervised Defect Segmentation by Applying Structural Similarity To Autoencoders

Paul Bergmann¹, Sindy Löwe^{1,2}, Michael Fauser¹, David Sattlegger¹, and Carsten Steger¹

¹*MVTec Software GmbH*

www.mvtec.com

{bergmannp, fauser, sattlegger, steger}@mvtec.com

²*University of Amsterdam*

sindy.loewe@student.uva.nl

Abstract—Convolutional autoencoders have emerged as popular methods for unsupervised defect segmentation on image data. Most commonly, this task is performed by thresholding a per-pixel reconstruction error based on an ℓ^p -distance. This procedure, however, leads to large residuals whenever the reconstruction includes slight localization inaccuracies around edges. It also fails to reveal defective regions that have been visually altered when intensity values stay roughly consistent. We show that these problems prevent these approaches from being applied to complex real-world scenarios and that they cannot be easily avoided by employing more elaborate architectures such as variational or feature matching autoencoders. We propose to use a perceptual loss function based on structural similarity that examines inter-dependencies between local image regions, taking into account luminance, contrast, and structural information, instead of simply comparing single pixel values. It achieves significant performance gains on a challenging real-world dataset of nanofibrous materials and a novel dataset of two woven fabrics over state-of-the-art approaches for unsupervised defect segmentation that use per-pixel reconstruction error metrics.

1. INTRODUCTION

Visual inspection is essential in industrial manufacturing to ensure high production quality and high cost efficiency by quickly discarding defective parts. Since manual inspection by humans is slow, expensive, and error-prone, the use of fully automated computer vision systems is becoming increasingly popular. Supervised methods, where the system learns how to segment defective regions by training on both defective and non-defective samples, are commonly used. However, they involve a large effort to annotate data and all possible defect types need to be known beforehand. Furthermore, in some production processes, the scrap rate might be too small to produce a sufficient number of defective samples for training, especially for data-hungry deep learning models.

In this work, we focus on unsupervised defect segmentation for visual inspection. The goal is to segment defective regions in images after having trained exclusively on non-defective samples. It has been shown that architec-

tures based on convolutional neural networks (CNNs) such as autoencoders (Goodfellow et al., 2016) or generative adversarial networks (GANs; Goodfellow et al., 2014) can be used for this task. We provide a brief overview of such methods in Section 2. These models try to reconstruct their inputs in the presence of certain constraints such as a bottleneck and thereby manage to capture the essence of high-dimensional data (e.g., images) in a lower-dimensional space. It is assumed that anomalies in the test data deviate from the training data manifold and the model is unable to reproduce them. As a result, large reconstruction errors indicate defects. Typically, the error measure that is employed is a per-pixel ℓ^p -distance, which is an ad-hoc choice made for the sake of simplicity and speed. However, these measures yield high residuals in locations where the reconstruction is only slightly inaccurate, e.g., due to small localization imprecisions of edges. They also fail to detect structural differences between the input and reconstructed images when the respective pixels' color values are roughly consistent. We show that this limits the usefulness of such methods when employed in complex real-world scenarios.

To alleviate the aforementioned problems, we propose to measure reconstruction accuracy using the structural similarity (SSIM) metric (Wang et al., 2004). SSIM is a distance measure designed to capture perceptual similarity that is less sensitive to edge alignment and gives importance to salient differences between input and reconstruction. It captures inter-dependencies between local pixel regions that are disregarded by the current state-of-the-art unsupervised defect segmentation methods based on autoencoders with per-pixel losses. We evaluate the performance gains obtained by employing SSIM as a loss function on two real-world industrial inspection datasets and demonstrate significant performance gains over per-pixel approaches. Figure 1 demonstrates the advantage of perceptual loss functions over a per-pixel ℓ^2 -loss on the NanoTWICE dataset of nanofibrous materials (Carrera et al., 2017). While both autoencoders alter the reconstruction in defective regions, only the residual map of the SSIM autoencoder allows a segmentation of these areas. By changing the loss function and otherwise keeping the same autoencoding architecture, we reach a performance that is on par with other state-of-the-art