Deep Learning Architectures for Detecting Surface Defects using Multiple illuminations

D. Honzátko, E. Türetken, S. Bigdeli, M. Kozinski •, P. Fua •, L. A. Dunbar

Application of deep-learning-based computer algorithms for surface defect detection is often problematic due to their need for large annotated datasets. Here we tackle this problem using a novel algorithm architecture that benefits from seeing the defect from different vantage points, i.e., multiple illuminations. Using self-supervised techniques, we capture defining, illumination-invariant, features of the surface which are later used for defect detection with a very limited annotated dataset.

Defect detection has long been of key importance to a wide range of industrial applications. Many of these applications are complex, require high-speed and yet must be cost-effective. Automated visual quality checks at frequent intervals have emerged as an efficient alternative to traditional inspection by (human) experts. In recent years, as with many other image analysis tasks, machine learning (ML) has become a standard tool used for defect detection.

The visibility of a defect depends on many factors such as illumination, reflection properties of the material, the relative orientation of the camera, etc. Nevertheless, to spot most of the defects, it is often sufficient to observe the surface using a light dome, where the camera is located on the top and the illumination angle can be controlled. Figure 1 shows an image of a watch part with many scratches captured using such a setup.

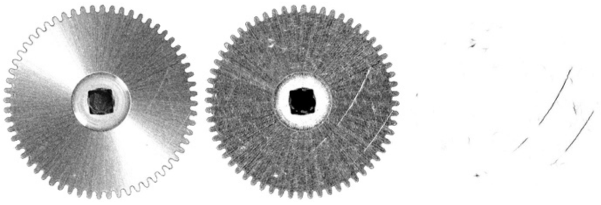

Figure 1: Defect detection of a watch part. (left) One of the many input images, (middle) Extracted features; (right) Defect predictions.

The manufacturer is usually interested in a binary decision about the defectiveness of a part. Nevertheless, due to the scarcity of the defects and resulting small number of samples that can be used to train and test ML algorithms, annotations have to be more precise – ideally, pixel-level-precise. Although the precise localization of defects dramatically reduces the problem of overfitting, the state-of-the-art deep learning architectures, such as the U-Net [1], still struggle to generalize well on relevant test sets. This can be attributed to the *curse of dimensionality* as each sample consist of multiple high-resolution images.

A well-known technique to tackle the *curse of dimensionality* is to first extract defining features from the sample and then use the ML algorithm on these features. For most image analysis tasks, the feature extraction part is performed by networks pretrained on large databases of natural images. Unfortunately, images of surfaces vary substantially in these databases and as such a more sophisticated solution is needed.

Here we propose a self-supervised feature extractor based on auto-encoder (AE) architecture. AE consists of two networks – Encoder, which extracts the features, and Decoder, which uses the features to estimate the original input. Although simple AE already exhibits minor improvements, further improvement by exploiting certain prior knowledge stemming from the controlled environment of a light dome may be possible. Such knowledge consists of the fact, that the image can be decomposed into object and light features where only the object features are useful for the defect detection.

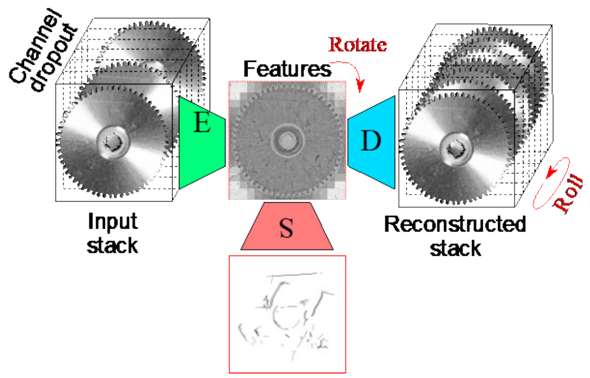

Figure 2: Rotate & Roll Channel Auto-Encoder architecture: Encoder (E) extracts the features, Decoder (D) estimates the original images, and Detector (S) segments the defects.

Such behavior can be imposed by Rotate & Roll Channel Auto-Encoder (RRCAE), which is visualized in Figure 2. It randomly rotates the features (before the decoding) as well as the desired image estimate. The desired output is also reordered (rolled) such that it seems to be illuminated from the same angle as if no rotation occurred. The random rotations are therefore limited to the angles of the possible illuminations. Such training process forces the features to contain object information only, while the light information gets implicitly encoded in the weights of the encoder and decoder network.

The core parts of the channel RRCAE – the Encoder, Decoder, and Detector are U-Nets, differing only in certain parameters, such as the number of output channels, up-sampling function, or the number of root features. To increase the generalization, during the training time, half of the input images are randomly dropped out (*i.e.*, multiplied by zero).

The proposed architecture was empirically tested on a dataset of 40 cogged wheels of size 512 x 512 px captured in a light dome under 12 different illuminations. Results of cross-validation on 20 defective samples shows that the defect detection performance of the model significantly improved compared to the standard U-Net solution. The mean AUROC score increased from 0.94 to 0.97 while the standard deviation dropped from 0.05 to 0.03.

As well as improving the defect detection with a reduced number of samples compared to the standard technique, as shown in Figure 1, this method provides a clear visualization of the intermediate features of the sample in a single image.

• CVLAB, EPFL, Lausanne

[1] O. Ronneberger, P. Fischer, T. Brox, “U-net: Convolutional networks for biomedical image segmentation“, MICCAI (2015).