Proceedings of Machine Learning Research 154:104–117, 2021 LIDTA 2021

GanoDIP - GAN Anomaly Detection through Intermediate Patches: a PCBA Manufacturing Case

Arnaud Bougaham ARNAUD.BOUGAHAM@UNAMUR.BE
Adrien Bibal ADRIEN.BIBAL@UNAMUR.BE
Isabelle Linden ISABELLE.LINDEN@UNAMUR.BE
Benoît Frénay BENOIT.FRENAY@UNAMUR.BE
Université de Namur - NADI - Belgium

Editors: Nuno Moniz, Paula Branco, Luís Torgo, Nathalie Japkowicz, Michał Woźniak and Shuo Wang.

Abstract

Industry 4.0 and recent deep learning progress make it possible to solve problems that traditional methods could not. This is the case for anomaly detection that received a particular attention from the machine learning community, and resulted in a use of generative adversarial networks (GANs). In this work, we propose to use intermediate patches for the inference step, after a WGAN training procedure suitable for highly imbalanced datasets, to make the anomaly detection possible on full size Printed Circuit Board Assembly (PCBA) images. We therefore show that our technique can be used to support or replace actual industrial image processing algorithms, as well as to avoid a waste of time for industries.

Keywords: Industry 4.0, AOI, PCBA, Anomaly Detection, Imbalanced Dataset, WGAN, Image Processing, Real-World Dataset, Unsupervised Learning

1. Introduction

In the last few decades, the industrial sector evolved with technologies, entering different successive revolutions. From the steam-powered equipment, to the introduction of electricity and IT equipment, the sector takes nowadays advantage of cyber-physical systems. Companies are now facing the 4th industrial revolution, also called Industry 4.0. This new paradigm involves a variety of key enablers, composed of the internet of things (IOT), analytics, data science, machine learning and decision systems. The main objective of these technologies is to optimize the factories productivity. In this context, deep learning, applied to automatic image inspection, offers a high potential to enforce quality control requirements. This can be done because factories own countless images that may be exploited, specifically to detect and localize anomalies in products.

This work proposes to solve a real-world, industrial, anomaly detection problem. Real-world images are considered, taken from the production lines of an Automatic Transmission Electronic Control Unit (ATECU) manufacturer. The product line is composed of successive processes, devoted to manufacture electronic Printed Circuit Board Assembly (PCBA), in order to equip car speed boxes. The first process is the one of interest in our work. It takes images of 100% of the products being manufactured through an Automatic Optical Inspection (AOI), applied on the 2 faces of the PCBA. To detect PCBAs with anomalies in the product line, traditional anomaly detection algorithms (comparison between an image under test and a golden sample image) are currently used in factories, through this process.

However, traditional anomaly detection algorithms have several practical drawbacks, which result on sub-optimal manufacturing operations. One drawback is a high false positive rate (false alarms), which involves useless manual checks of the PCBAs. In order to solve this industrial problem, the machine learning literature on anomaly detection can be used to reduce the pseudo-anomalies. However, the datasets used in the literature do not always represent the real industrial world, which requires some tuning for techniques to be suitable.

This work is motivated by the poor performance of state-of-the-art techniques on our industrial problem. Yet, GAN-based methods are promising to address this issue, specifically f-AnoGAN (Schlegl et al., 2019). This method is an improvement of the first GAN-based solution for anomaly detection based on real-world images (optical coherence tomography imaging data of the retina). We build on f-AnoGAN and propose several adaptations that help solving our industrial problem.

The main contribution of this work is a modification of the state-of-the-art f-AnoGAN, to deal with our real-world industrial problem. In contrast to this method, we consider full images as an input, with patch decomposition at the inference step to better localize anomalies. This will greatly reduce the amount of products with pseudo-anomalies sent to the human operator for inspection.

This work is structured as follows. Section 2 first defines the problem and the industrial issues. Section 3 details our real-world industrial dataset. Section 4 reviews the relevant techniques in the literature that can be used to tackle our problem. In order to use them in a real-world industrial setting, Section 5 proposes adaptations to f-AnoGAN, quantitatively and qualitatively evaluated in Section 6. Finally, Section 7 concludes the paper.

2. Problem Definition and Industrial Issues

Automatic Transmission Electronic Control Units (ATECUs) are composed of a Printed Circuit Board Assembly (PCBA) inserted in a sealed case, intended to be mounted and connected to a car speed box, before being sent to a car manufacturer. Many complex operations are required to achieve this goal, with high volume and quality constraints driven by the safety importance an automotive product requires.

In this work, the very first stage of this PCBA manufacturing is considered, namely the visual test at the Automatic Optical Inspection (AOI) process, right after the components mounting and soldering onto the blank PCB. It is commonly agreed that the earlier a quality issue is detected, the best it can be counter-measured and contained.

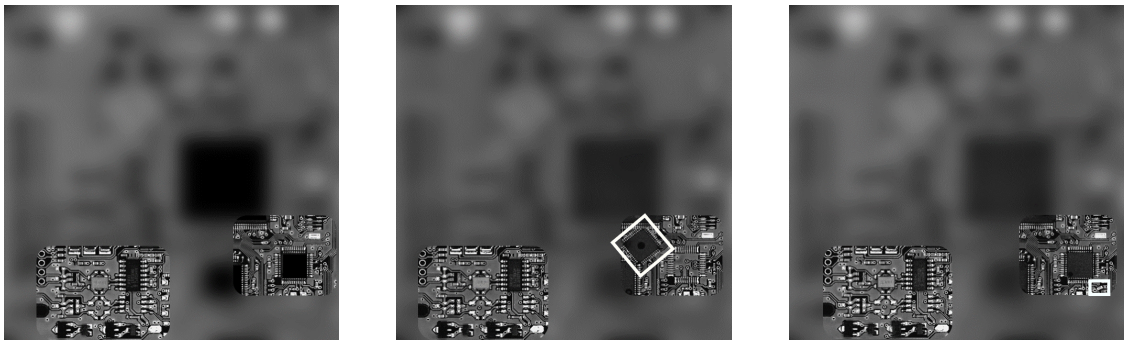
The complexity of the operations prior to the AOI makes it challenging to achieve the quality requirements, and anomalies can be produced in many ways. The role of the AOI process is to reject PCBAs with visible anomalies. In order to do so, a camera takes high definition images of the product being conveyed, and an traditional image processing algorithm is in charge of detecting anomalies in the product.

The AOI process is a key element of the manufacturing line, and has to operate with the highest confidence degree. The image processing algorithm is designed to ensure no missed detection (type II error), which involves severe test limits and generates many false alarms (type I error). When a pseudo-anomaly is raised by the algorithm, a human operator is asked to confirm or infirm the detection by visually scanning the area of concern.

This visual anomaly judgment guarantees the product quality, but is time-consuming (±8s per PCBA inspection) and prone to human error. As a consequence, this manual operation slows down the line throughput, increases internal costs and quality risks (the more the pseudo-anomalies are false positives, the more probable it is to make an error on true positives), which highly participates in indirect global wastes. However, it is not conceivable to reduce the false positive rate (FPR) by decreasing the test limits severity, at the risk of increasing the false negative rate (FNR). This would be equivalent to accept the risk of manufacturing products with anomalies. This process is therefore continuously monitored and parameters are set to reduce the FPR, while keeping the lowest FNR possible.

3. High Resolution and Small Details: the PCBA Dataset

As explained in Section 2, images from the AOI camera are processed to assess the product quality. Our imbalanced dataset contains 428 grayscale JPEG images: 410 normal (without anomalies) and 18 abnormal (with anomalies) images. Each $4,500 \times 4,340$ image has undergone an adjusted xy re-orientation (via fiducial reference centering) during acquisition. PCBA images reveal some specific characteristics that have to be considered for our method. In particular, our images are of high-resolution and have anomalies residing in details. For instance, one can see in Figure 1(a)[1] that many information is contained in the overall image.



(a) normal (b) large anomaly (c) small anomaly

Figure 1: Figure 1(a) presents a normal PCBA. On the other ones, the white frames surround a large anomaly (Figure 1(b)) and a small one (Figure 1(c)). Some parts of the images have been anonymized (material under intellectual property).

Some areas of the PCBA images do not contain much information, while some others present many details. Non-informative areas show progressive changes in pixel intensity, whereas components-dense areas yield to abrupt changes in pixel intensity. This is a difficulty that the anomaly detection method has to deal with. Anomalies occur in limited parts of the image. They can result in important changes (e.g., absence of component) or very tiny changes (e.g., bridge on integrated circuit pins). Anomaly detection is particu-

1. Some parts of the images have been blurred to guarantee the intellectual property of our industrial partner. The arguments described also apply for the hidden parts, where information can be extrapolated.

larly tricky for small detailed changes, in particular for dense areas with many components. Therefore, it is important to have images of sufficient resolution that reveal these details.

Since this work aims to differentiate normal and abnormal images, it is interesting to illustrate how they can differ. Figure 1 shows two very different anomalies, Figure 1(b) being a large anomaly, while Figure 1(c) being a smaller one. One can see how difficult it is to figure out the small defect in this overcrowded image. Small defects cause normal and abnormal images to be hard to distinguish for state-of-the-art techniques. Besides, even for normal images, there is a locally high pixel intensity variability that is also challenging.

4. State-of-the-Art Anomaly Detection Techniques

Usual techniques for detecting anomalies are unsupervised, given that abnormal labeled data are almost non-existent in many contexts, including our industrial one (Ruff et al., 2021). In addition to this highly imbalanced dataset constraint, it is possible that the few available detected and labeled abnormal data do not represent the entire spectrum of possible anomalies.

Variants of classical models have first been considered for this task. One-class classification techniques, such as One-Class Support Vector Machine (OC-SVM) (Schölkopf et al., 2001) and Support Vector Data Description (SVDD) (Tax and Duin, 2004) have been developed for this purpose. Reconstruction approaches like Kernel Principal Component Analysis (Kernel PCA) have also been considered to solve the task (Hoffmann, 2007), outperforming OC-SVM. Nevertheless, even if these methods are efficient for low-dimensional data, a poor discriminative performance has been observed with high-dimensional images (Cheng et al., 2020). Due to this limitation, deep learning variants have been preferred, resulting in Deep OC-SVM (Erfani et al., 2016) and Deep SVDD (Ruff et al., 2018).

Autoencoders (AE) (Hinton and Salakhutdinov, 2006) (and all of its variants) have also gained much attention. They bring better performance when reconstructing normal data, yet they also result in abnormal data that are perfectly reconstructed, *including defects*. This is particularly observed on complex data with inherent small normal variations, which lead to confusion with small defects. This is a key drawback, as it reduces their discriminative capacity, based on a comparison of original and reconstructed images.

Adversarial learning-based models, in the context of anomaly detection, have been developed to tackle this issue. The idea of Generative Adversarial Networks (GANs) is to adversarially train a generator and a discriminator, so that the generator captures, as faithfully as possible, the distribution of the normal data. If the discriminator is not able to distinguish a real image from a generated image, the generator is considered efficient enough to be the decoder part (Goodfellow et al., 2014).

A restricted latent space developed by Perera et al. (2019) yields to an input image representation always similar to the one for the class trained. The key elements of their OCGAN method lie in (i) minimizing the MSE between the original and reconstructed images, (ii) constraining a latent discriminator to be uniformly distributed (to avoid out-of-class representation possibilities), and (iii) ensuring the generated image realness thanks to a visual discriminator. The generator and the discriminators are all adversarially trained.

GAN training in f-AnoGAN is performed in two steps (Schlegl et al., 2019). The first one is the decoder and discriminator training, the second one is the encoder training. At

inference, the anomaly score is composed of the mean squared error (MSE) and the discriminator loss between the original and reconstructed images. This method received a great attention, since it was the first attempt to use GANs for anomaly detection. Unlike other techniques, f-AnoGAN (Schlegl et al., 2019) uses real-world images, i.e. it is able to deal with complex datasets, like our PCBA images. This method is detailed in Section 5.1.

Since f-AnoGAN trains a GAN with random patches of the input image, the associated latent space is not designed to reconstruct a real, full image. This issue, identified by the authors of that technique, has a real practical importance in industry. Indeed, showing the reconstructed image to users can improve the trust that they put in the method and, thus, accelerate its adoption. All the more, in our case, the location and visual context of changes in details is important to assess whether it corresponds to an anomaly or not. The input image must therefore be considered as a whole, which is elaborated in the next section.

5. GAN Anomaly Detection through Intermediate Patches: GanoDIP

The proposed method, called GanoDIP for GAN Anomaly Detection through Intermediate Patches, is a modified f-AnoGAN (Schlegl et al., 2019) whose inference step has been adapted. Indeed, the changes are designed for the needs of the analysis of high-resolution images where defects can correspond to small details, whose locations in the input image is important. Both training and inference steps are detailed in this section.

5.1. Training Step of f-AnoGAN

This f-AnoGAN training step has been selected for its ability to reconstruct normal images while not reconstructing abnormal ones (i.e., defects are not preserved in the reconstructed image). This key element for anomaly detection allows us to generate realistic images that are visually controllable by experts, in order to gain trust in the network performance. Figure 2 gives an overview of the overall training strategy.

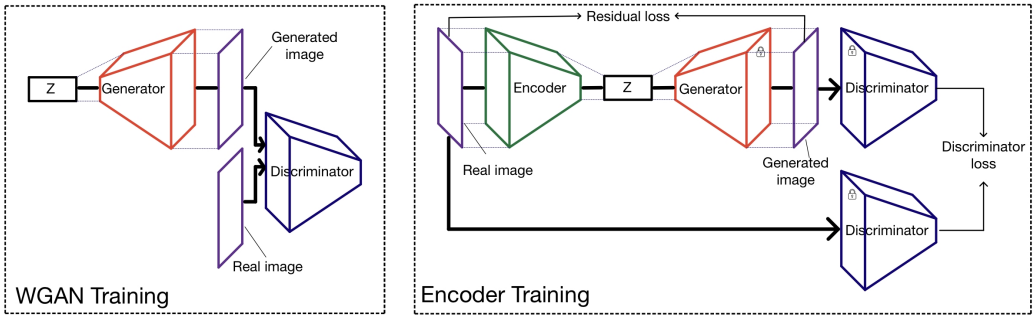


Figure 2: Figure inspired from Schlegl et al. (2019) presenting the training strategy of f-AnoGAN, which is used as the first step of the proposed method GanoDIP.

Two convolutional networks, being a generator G (decoder-based) and a discriminator D (encoder-based) are trained simultaneously with normal images. The generator is fed by a normally distributed latent vector of dimension 90. The discriminator alternatively receives

real and generated images, and calculates the distance between both. In order to stabilize the training process, the Wasserstein GAN (WGAN) (Arjovsky et al., 2017) is chosen with gradient penalty (Gulrajani et al., 2017), which allows us to evaluate the distance between the real distribution $\mathbb{P}_r$ and the generated distribution $\mathbb{P}_g$. The loss is described by:

$$\mathcal{L}_{GPWGAN} = \mathop{\mathbb{E}}_{\tilde{\mathbf{x}}\sim\mathbb{P}_g}[D(\tilde{\mathbf{x}})] - \mathop{\mathbb{E}}_{\mathbf{x}\sim\mathbb{P}_r}[D(\mathbf{x})] + \lambda \mathop{\mathbb{E}}_{\hat{\mathbf{x}}\sim\mathbb{P}_{\hat{\mathbf{x}}}}[(||\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}})||_2 - 1)^2], \tag{1}$$

where $\tilde{\mathbf{x}}$ is a sample decoded from a latent vector encoding, $\mathbf{x}$ is an original sample, $\hat{\mathbf{x}}$ is a sample generated from the generator G, $D(\cdot)$ is the discriminator and the right-hand term of the Equation (1) refers to the gradient penalty (Gulrajani et al., 2017).

The above loss function makes it possible to generate images that can fool the discriminator. Once the generator and the discriminator are trained, they remain frozen (i.e., their parameters are fixed) for the following steps.

An encoder E is then trained to map the images into the latent space. To do so, a traditional autoencoder architecture is used to encode/decode the training images, controlled by the MSE between the original and the reconstructed image (the residual loss). The previously trained decoder being frozen, only the encoder parameters are updated. The original and reconstructed images are also fed into the frozen discriminator, and the discriminator loss is calculated. The encoder loss is therefore driven by a combination of the residual loss (similarity in the image space) and the discriminator loss (similarity in the feature space):

$$\mathcal{L}_{encoder}(\mathbf{x}) = \frac{1}{n}||\mathbf{x} - G(E(\mathbf{x}))||^2 + \frac{\kappa}{n_d}||f(\mathbf{x}) - f(G(E(\mathbf{x})))||^2, \tag{2}$$

where n is the number of samples in the dataset, $G(E(\mathbf{x}))$ is the reconstructed image, $f(\cdot)$ is the transformation to the feature space, n_d is the number of dimensions of the feature space and κ balances the two terms (Schlegl et al., 2019).

The main limitation specified by the authors of f-AnoGAN (Schlegl et al., 2019) is the use of random 64×64 patches, which divides the original full-size image into smaller parts at the beginning. Even if the progressive pixel changes observed in their images reduce the context constraint, a Visual Turing Test (distinction between real and generated images shown to domain experts) can only be performed on non-standard small patches that experts are not used to work on. A similar test would not be achievable with our dataset, as the context has a great importance. Indeed, some areas can show the presence of elements, whereas some others can show a similar situation with no element, and still be completely normal. Unlike the images considered in the work of Jiang et al. (2019), where f-AnoGAN is used for the identification of missing cigarettes boxes, no matter their position, the location of the missing parts are important in the PCBA images. For instance, depending on the location, some components can be mounted or not on their lands. The Visual Turing Test validity is even more important when only a few number of abnormal images are available to quantitatively evaluate the method, which is often the case in anomaly detection. This Visual Turing test, among other qualitative ones, is of particular importance, bringing trust to the network ability to generate realistic images.

Our contribution tackles the above issues and is intended to deal with contextually rich datasets composed, as our images, of a combination of smooth pixel intensity variations (e.g., on solder pads) and abrupt pixel intensity variations (e.g., on component locations).

5.2. Inference Step of the Proposed Method GanoDIP

The proposed GanoDIP method considers full-size input images for training. This leads to difficulties while comparing the reconstructed and the original images, due to the increased number of regions containing normal variations. As a consequence, the discrimination between normal and abnormal images becomes tedious. Unlike f-AnoGAN, a specific treatment is therefore required to tackle this difficulty.

The idea is to use patches at the inference step, so as to focus on small areas at this stage only, and to get rid of the global noise of the full-size image. The entire image is therefore fed into the autoencoder (composed of the encoder and the decoder previously trained) that generates a reconstructed image. Then, both the original and the reconstructed images are sliced into patches (dividing the images in equal parts), namely $\mathbf{p} = \{\mathrm{p}_0, \mathrm{p}_1, ..., \mathrm{p}_N\}$ for the original patches, and $\hat{\mathbf{p}} = \{\hat{\mathrm{p}}_0, \hat{\mathrm{p}}_1, ..., \hat{\mathrm{p}}_N\}$ for the reconstructed ones (N being the total number of patches). These small areas, called intermediate patches due to their presence after the training step but before the final inference result, are used to compute the MSE_i performed pixel-wise, for each p_i and $\hat{\mathrm{p}}_i$, i being the patch index.

Once all of these patch reconstruction losses are obtained, the worst $\alpha\%$ of these losses are averaged, where α is a hyperparameter. The aggregated patch reconstruction loss is the anomaly score associated to the queried image, i.e.,

$$\text{Anomaly Score} = \frac{1}{m}\sum_{k=1}^{m}\{\mathrm{MSE}_k \mid \mathrm{MSE}_k \in m \text{ highest } \mathrm{MSE}_i\}, \tag{3}$$

where m corresponds to the $\alpha\%$ of the patches with the highest anomaly score, and i is the patch index. This inference step is graphically summarized in Figure 3.

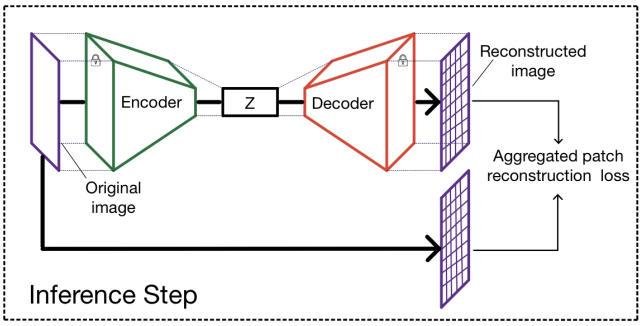


Figure 3: Inference strategy of our method GanoDIP. The entire image is fed into an autoencoder that reconstructs the image. The original and the reconstructed images are sliced into intermediate patches and used to compute the MSE performed patch by patch. The worst MSEs are then aggregated to give the reconstruction loss.

In addition to making it possible to use high-resolution/small details images that we encounter in our real-world industrial scenario, GanoDIP also enables to easily localize

anomalies. Indeed, the intermediate patches, used at the inference step, automatically give the location of the highest identified anomalies, while overlaying them on the original image. This gives a clear view on the areas difficult to reconstruct, indicating a probable anomaly.

6. Evaluation

This section evaluates the proposed methodology. The pre-processing of the dataset is detailed in Section 6.1, then the results of quantitative experiments are presented in Section 6.2 and confirmed by qualitative experiments reported in Section 6.3. All experiments have been undertaken with an RTX 2080 TI GPU, Python 3, Cuda 10 and Tensorflow 2.

6.1. Dataset Pre-Processing and Experimental Protocol

For our experiments, 360 normal images are used for the training of the WGAN and the encoder, while 50 normal and 18 abnormal images are reserved for testing. They all have undergone an adjusted xy re-orientation (via fiducial reference centering) during acquisition. Resizing and normalization pre-processing are applied before usage. The final image sizes in the experiments are 512×512 and 384×384. Smaller sizes were discarded by preliminary analysis, explained by the level of details and the small size of some anomalies in the images.

Because of the nature of our problem, only 18 anomaly samples are available, which is too small to propose any quantitative metric with enough statistical representation. For this reason, as this is classically done in the anomaly detection literature (see, e.g., Schlegl et al. (2019)), we only show the anomaly score distributions between the two classes (normal vs. abnormal). However, to go beyond that analysis, a threshold on the anomaly score is also computed to separate normal and abnormal test images as much as possible. Again, due the small number of the anomaly samples, this threshold could not be cross-validated, but corresponds to the lowest FPR when a zero FNR on the test set is required.

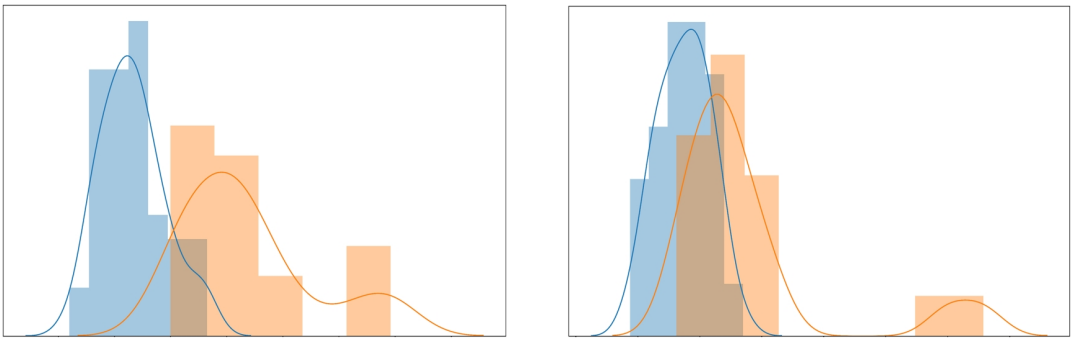
For our experiments, several hyperparameters must be set. Some of them were set after preliminary experiments: the GAN training epochs number is 750, the encoder training iterations number is 20,000. Below these numbers, underfitting is observed yielding to areas not well reconstructed, with visible artefacts and information leakage on the image. The percentage of intermediate patches to be kept is 0.15% and the intermediate patch size is 4×4. In the following experiments, two input image sizes (512×512 and 384×384) are considered to evaluate their relevance.

6.2. Quantitative Assessment

Based on the idea of focusing on the worst outliers to discriminate normal and abnormal images, the α=0.15% highest patch anomaly scores are averaged, to form the image anomaly score, as described in Equation (3). This image anomaly score is the metric that is taken in consideration to determine if a PCBA is normal or abnormal. The larger this anomaly score is, the more likely the PCBA is to be abnormal. It means that the discriminative capacity lies into the distinction between the anomaly score distributions of both classes.

Figure 4 presents the anomaly score distributions on a test set, where the normal and the abnormal image distributions are colored in blue and orange respectively. These distributions correspond to the frequency of anomaly scores for the two classes. This figure

shows the importance of the input image size choice, resulting in an FPR of 14% for the 512×512 configuration (Figure 4(a)) and a FPR of 60% for the 384×384 configuration (Figure 4(b)). This confirms the sensitivity to the input image size and is again explained by the fact that a higher resolution preserves smaller details. In practice, the reported FPR of 14% would decrease the inspection time down to 2.24 seconds per PCBA. This is a significant improvement from the 8-second inspection per PCBA observed in the industry, with the current techniques. This means that 72% of the inspection time would be saved for the human operator.



(a) 512×512 (b) 384×384

Figure 4: Anomaly scores for normal (blue) and abnormal (orange) images of the test set, with image sizes of 512×512 (Figure 4(a)) and 384×384 (Figure 4(b)). For each figure, the x-axis is the anomaly score and the y-axis is the score frequency.

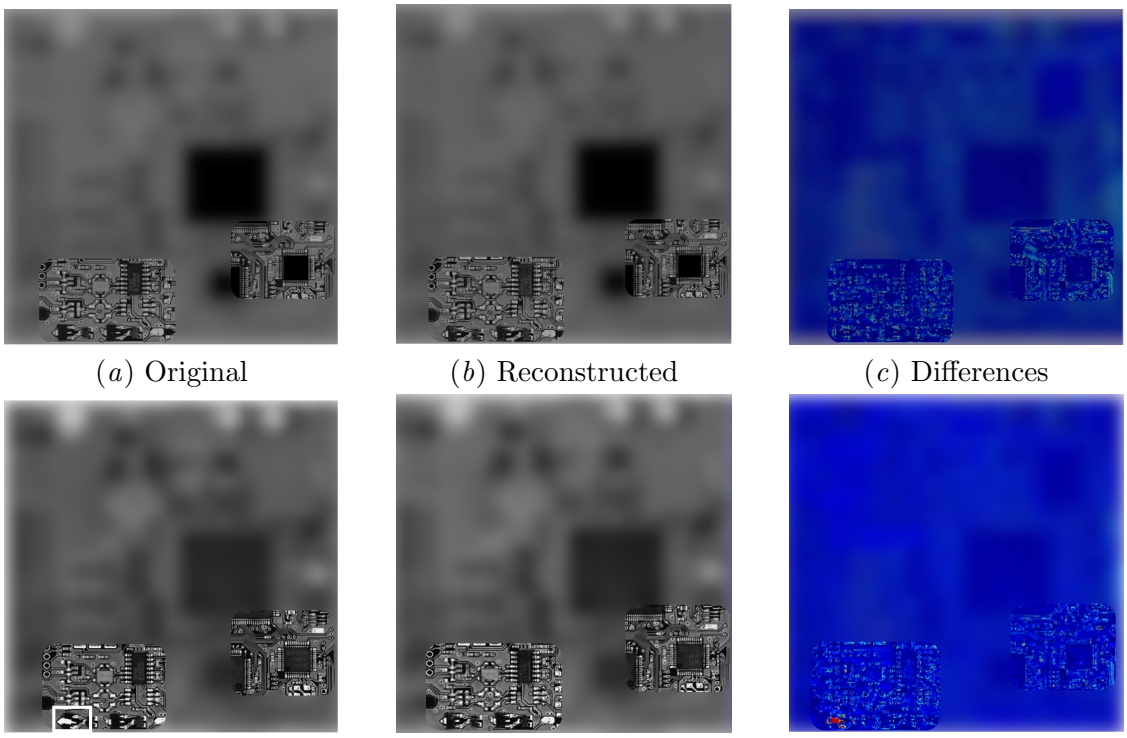
The quantitative assessment presented in this section demonstrates that the discriminative capacity between normal and abnormal images is driven by a few set of intermediate patches, the ones where the reconstructed image has significant differences with the original one. These intermediate patches can be used to identify the presence of anomalies, and have to be isolated. The isolation of these patches and the global anomaly score computed is key to our method. Moreover, this assessment shows the influence of the input image size, which also is a key element for preserving the small details characterizing some anomalies.

6.3. Qualitative Assessment

As the above quantitative experiment is based on a limited set of abnormal images, GanoDIP is hereafter also assessed qualitatively and with a Visual Turing Test. This is done with the best hyperparameters from the previous section, namely an input image size of 512×512.

The images shown in Figures 5(c) and 5(f) highlight the differences between the original (Figures 5(a) and 5(d)) and the reconstructed (Figures 5(b) and 5(e)) images. The pixel differences state whether an anomaly is present or not.

Regardless of presence or absence of an anomaly, the high majority of areas shows no difference between pixels (the blue ones), which means that the original image has been correctly reconstructed. Several incorrectly reconstructed areas (green and red pixels) in the top right image are actually normal areas. These areas are subject to high, but normal variations (e.g., different markings on components, solder luminosity, components slightly



(*a*) Original (*b*) Reconstructed (*c*) Differences

(*d*) Original (*e*) Reconstructed (*f*) Differences

Figure 5: Original (Figures 5(*a*) and 5(*d*)), reconstructed (Figures 5(*b*) and 5(*e*)) and difference between both (Figures 5(*c*) and 5(*f*)), for a normal (first row) and an abnormal image (second row). The anomaly in Figure 5(*d*) is white-framed for a better identification. Some parts of the images have been anonymized (material under intellectual property).

shifted when soldered, etc.). In the bottom right image of Figure 5, clusters of red pixels can be spotted, making the anomaly detection possible. It can also be seen in the middle column of Figure 5 that the images decoded by the WGAN are realistic and do not preserve the anomaly (if any). This is key in our context, as they can be used as a proof that GanoDIP is working well, in order to gain the trust of people in the industry. As mentioned before, this is a major short-coming of f-AnoGAN, which can only generate parts of the global image (see the future work in (Schlegl et al., 2019)).

The difficulty in discriminating high variations in normal areas and in abnormal areas is directly related to complexity of the images in our dataset. Indeed, because of the high-resolution of our images and of the small details in anomalies, normal and abnormal images look alike for the state-of-the-art techniques. These elements allow us to deduce that the intermediate patches are necessary to isolate the highest anomaly areas in the full size image. These intermediate patches make it possible to discard the high, but normal, variations.

Regarding the previously stated difficulty and anomaly localization concern, Figure 6 shows the original image with superimposed intermediate patches for three abnormal PCBAs.

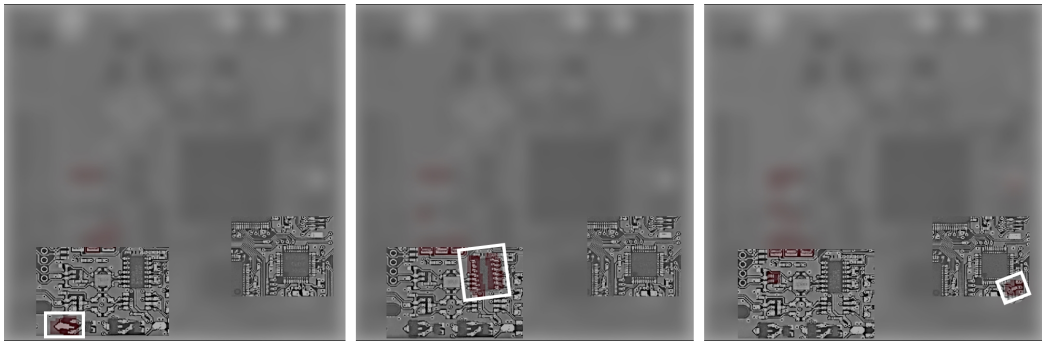


Figure 6: Three original abnormal images with intermediate patches (red areas) used to localize the anomalies. All anomalies on the PCBAs are well identified and localized (white framed for the ease of the reader). Some parts of the images have been anonymized (material under intellectual property).

This superimposition makes it possible to reveal the areas that are difficult to reconstruct. Unlike Figure 5, anomaly locations are isolated by taking the anomalies present in the worst intermediate patches. Thanks to them (red pixels patches), a human operator asked to check the image would only spend few seconds to figure out that a defect is present on the PCBA, which is a considerable improvement over the difference image of Figure 5. This gives a clearer view on where he/she should focus his/her attention, when asked to confirm or infirm the defects.

It should be mentioned that the variability of the reconstructed images, while decoding the original images, is very low, due to the low dispersion of the original full size images, giving an overfitting-like behavior. When randomly sampling the latent vector within the image domain, the reconstructed images barely vary, which results in similar images being used as a reference for the inference step. An argument raised by this observation can be that a concurrent and easier solution exists: select a single, golden sample in the training dataset, and use it as a reference instead of the decoded image of GanoDIP. To prove the added value of our method with respect to this golden sample strategy, 36 training images have been randomly sampled to compose 10 new datasets (our original training set contains 360 images). Then, for each datasets, we calculated the FPR while taking each image as a reference, and selected the one with the smallest FPR as a golden reference. These 10 references are used to state how they can classify normal and abnormal images in the test dataset, for any random input dataset.

The FPR comparison for both the golden sample and the reconstructed strategies is shown in Figure 7. This figure demonstrates how unlikely it is to choose a good reference training image, while GanoDIP can build it for all training datasets. Indeed, most of the best golden samples fail to achieve results that compete with GanoDIP. Taking a sample of reference in the training dataset almost always provides distributions that are difficult to discriminate. This means that even if our method produces reconstructed images that are somewhat similar to each other, it can solve the problem for any dataset with the same characteristics as ours, while the golden sample strategy only seems to work in specific cases.

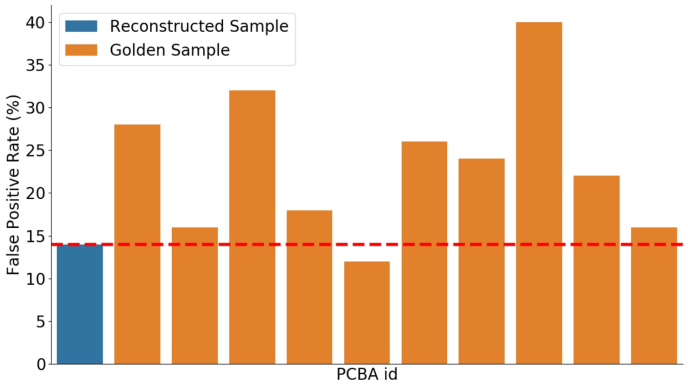


Figure 7: False positive rate (FPR) comparison between the reconstructed and the golden sample strategies. Our FPR (in blue) of 14% is compared to the golden sample having the lowest FPR in 10 groups of 36 random training images (in orange).

The quality of the reconstructed images being highly important in anomaly detection, a Visual Turing Test (VTT) has been performed with 5 domain experts with different expertise, all working with PCBAs. This test evaluates the realness of reconstructed images, which reflects how the latent vector fed into the decoder yields realistic images, and, thus, will be able to provide the best reference possible at the inference step. A classic VTT protocol (see, e.g., Schlegl et al. (2019); Han et al. (2018)) has been followed, initially proposed by Salimans et al. (2016). Latent vector values are randomly sampled to generate 50 images. These generated images and 50 original ones are randomly presented one by one to a group of participants, who have 16 seconds to label their realness (original or generated). Between each image, nothing is presented during 5 seconds, to reset the visual memory of the participants and to make them ready to focus on the next image. No indication is given about how many images are in each class and no training on the distinction between the original and generated images is provided before the evaluation. A 5 minutes break is given at the middle of the sequence, with no possibility for them to talk. An example of a sequence of two images is shown at the beginning, without any explanation, to understand how images are presented during the evaluation. A timer and the image id are shown.

The results of the VTT is an accuracy of 58.2% of the images being correctly labeled. This demonstrates the realness of the generated images and their adequateness as references for the inference step. Indeed, experts seem to randomly guess how to detect the generated PCBA, showing the difficulty to distinguish real images from generated ones, and thus proving their similarity.

This section shows that the ability to discriminate between normal and abnormal images is driven by an important set of intermediate patches. The use of intermediate patches, after the encoder-decoder step, enables the isolation of the anomalies, and makes it possible to localize them on the original full-size image. Furthermore, the FPR is greatly reduced compared to the image processing algorithm currently used. From a business point of view, this achievement is considered as a significant improvement. The time spent is not only decreased by the FPR reduction, but also by the fact that clear areas are provided for the

operator to pay attention. It should also be noted that the inference step proposed with our method GanoDIP is roughly 85% faster than the image processing non-machine-learning-algorithm actually used by our industrial partner. The main limitation lies in the abnormal images scarcity, i.e., quantitative results should be considered with care.

7. Conclusion

In this work, a method is proposed to tackle an anomaly detection task on a real-world industrial dataset, being highly imbalanced. No matter the skewness degree of the dataset, the one-class nature of the learning technique enables a full applicability on all imbalanced problem. As state-of-the-art techniques are not suitable to tackle our problem, we propose to extend f-AnoGAN with intermediate patches (GanoDIP) to make it usable on high-resolution images that may contain small or large anomalies, and whose localisation in the image is important.

The industrial context adds constraints. For instance, images in the latent space need to be realistic, so that they can be shown to domain experts to gain their trust in the generative model, and thus its adoption. It is also required to minimize the false positive rate under a zero-defect constraint, instead of optimizing any accuracy, precision or recall metric, as classically done.

Our experiments show that our proposed method can deal with the above challenges. A false positive rate of 14% can be achieved, resulting in 2.24 seconds of manual inspection time, which is far below the 8 seconds currently observed. This means that 72% of the time could be saved to check products by a human operator, if our method is deployed.

These promising results raise new questions to be investigated in future works. In particular, it would be interesting to study the behaviour of our method when trained for multiple product lines. Indeed, the currently considered dataset comes from a particular product line, with specificity cameras (e.g., in terms of luminosity). In practice, a new model would be needed for each product line, motivating future work to create a single model for all product lines in a factory.

References

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 214–223, 2017.

Haoqing Cheng, Heng Liu, Fei Gao, and Zhuo Chen. ADGAN: A scalable GAN-based architecture for image anomaly detection. In *Proceedings of the IEEE Information Technology, Networking, Electronic and Automation Control Conference (ITNEC)*, volume 1, pages 987–993, 2020.

Sarah M Erfani, Sutharshan Rajasegarar, Shanika Karunasekera, and Christopher Leckie. High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning. *Pattern Recognition*, 58:121–134, 2016.

Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Proceedings*

of the International Conference on Neural Information Processing Systems (NeurIPS), 2014.

Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of Wasserstein GANs. In *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2017.

Changhee Han, Hideaki Hayashi, Leonardo Rundo, Ryosuke Araki, Wataru Shimoda, Shinichi Muramatsu, Yujiro Furukawa, Giancarlo Mauri, and Hideki Nakayama. Gan-based synthetic brain mr image generation. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 734–738, 2018.

Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

Heiko Hoffmann. Kernel PCA for novelty detection. *Pattern Recognition*, 40(3):863–874, 2007.

Yu Jiang, Wei Wang, and Chunhui Zhao. A machine vision-based realtime anomaly detection method for industrial products using deep learning. In *Chinese Automation Congress (CAC)*, pages 4842–4847, 2019.

Pramuditha Perera, Ramesh Nallapati, and Bing Xiang. OCGAN: One-class novelty detection using GANs with constrained latent representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2898–2906, 2019.

Lukas Ruff, Robert Vandermeulen, Nico Goernitz, Lucas Deecke, Shoaib Ahmed Siddiqui, Alexander Binder, Emmanuel Müller, and Marius Kloft. Deep one-class classification. In *International Conference on Machine Learning (ICML)*, pages 4393–4402, 2018.

Lukas Ruff, Jacob R Kauffmann, Robert A Vandermeulen, Grégoire Montavon, Wojciech Samek, Marius Kloft, Thomas G Dietterich, and Klaus-Robert Müller. A unifying review of deep and shallow anomaly detection. *Proceedings of the IEEE*, 2021.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training GANs. *Proceedings of the International Conference on Neural Information Processing Systems (NeurIPS)*, 2016.

Thomas Schlegl, Philipp Seeböck, Sebastian M Waldstein, Georg Langs, and Ursula Schmidt-Erfurth. f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks. *Medical Image Analysis*, 54:30–44, 2019.

Bernhard Schölkopf, John C Platt, John Shawe-Taylor, Alex J Smola, and Robert C Williamson. Estimating the support of a high-dimensional distribution. *Neural Computation*, 13(7):1443–1471, 2001.

David MJ Tax and Robert PW Duin. Support vector data description. *Machine Learning*, 54(1):45–66, 2004.