Comparing discriminative and generative approach to detection of defects on injection moulded routing disc

Samo Šela[1,2], Domen Tabernik[1], Jan Ogrin[2], Danijel Skočaj[1]

[1] *University of Ljubljana, Faculty of Computer and Information Science, Večna pot 113, 1000 Ljubljana, Slovenia*
[2] *Kolektor Group d. o. o., Vojkova 10, 5280 Idrija, Slovenia*

E-mail: samo.sela@kolektor.com, domen.tabernik@fri.uni-lj.si, jan.ogrin@kolektor.com, danijel.skocaj@fri.uni-lj.si

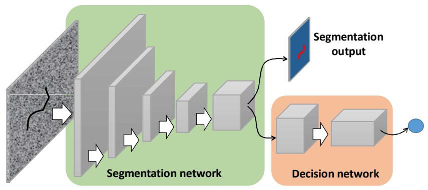
Abstract

A key part of zero-defect manufacturing is automatic image-based detection of defects in products and materials. In this paper, we explore two distinctive state-of-the-art approaches for defects detection: (a) fully discriminative approach, and (b) generative approach. Although discriminative approaches are currently achieving state-of-the-art results, their application is limited to domains with an abundant number of faulty samples. On the other hand, generative approaches minimize the use of faulty samples and can learn the representation only from defect-free samples, however, their classification performance seems still to be limited. In this paper, we compare the representatives of discriminative (DecisionNet) and generative (GANomaly) approaches on a selected industrial domain of injection-moulded-parts inspection. We show that the generative method is still outperformed by the discriminative approach with mAP of 99.3% against 99.8%, however, the generative method is comparable when measuring their performance on a per physical-instance basis.

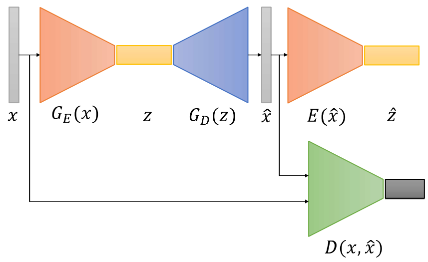
1 Introduction

In smart factories of the future, production can range from a highly individualized, batch-size-one production to a single mass-produced product under the conditions of high flexibility [6]. To achieve zero-defect manufacturing in this new paradigm, the methods of automatic quality control must be adapted as well. Lately, machine learning and deep-learning approaches have been showing significant promises in addressing this issue.

However, general and completely automatic recognition of anomalous products during the manufacturing process is still an unsolved problem. Understanding current operating conditions and detecting faults and failures is an important research topic. Complex visual inspection tasks are often done by human workers. Such manual labor is a slow and cumbersome process, and it is prone to errors. Replacing this process with an automated computer vision approach requires addressing two key issues arising from the nature of the production process: (a) high variability of the anomaly appearance and (b) highly unbalanced datasets with many defect-free parts but only a few faulty ones available.



(a) DecisionNet



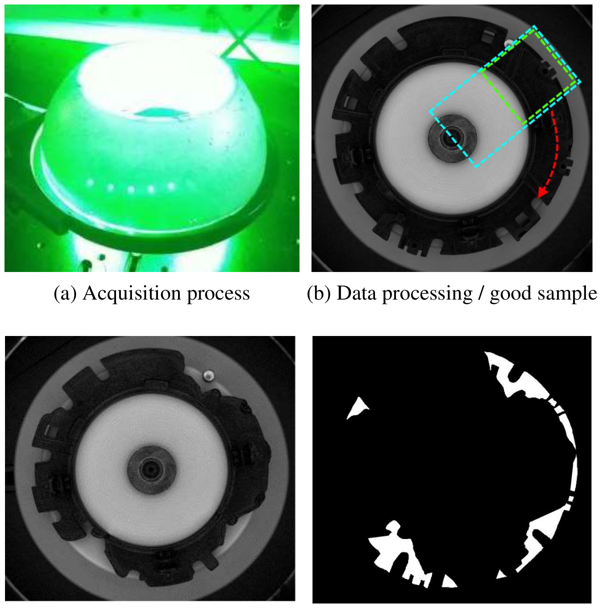
(b) GANomaly

Figure 1: (a) DecisionNet is a two-stage discriminative deep neural network architecture [5]. (b) GANomaly is a conditional GAN with an encoder-decoder-encoder subnetworks architecture [1].

Works by [7, 3, 5] have addressed the high variability of the anomaly appearance using deep-learning approaches, but their focus has been on fully supervised learning. To correctly model anomalies, many examples that capture different variations of the defects are required. When only a small number of training samples are available, the classification performance decreases, as shown by Tabernik el al. [5]. On the other hand, methods by [4, 8, 1] use Generative Adversarial Networks (GAN) [2] to learn the core anomaly detection model from only defect-free samples. This addresses the issue of highly unbalanced datasets, with only a minor number of faulty samples required to calibrate the detection process.

In this paper, we compare a discriminative approach against a generative approach on a real-world industrial case of anomaly detection. As a discriminative approach, we employ a two-stage deep-learning architecture with the segmentation network in the first stage and the decision network in the second stage [5], where both models are learned in a fully discriminative manner using pixel-wise labels of the defects. In this paper, we refer to this network as DecisionNet. As a generative approach, we utilize a generative adversarial network with an encoder-decoder-encoder structure termed GANomaly [1]. This architecture mitigates the problem in highly biased datasets (many defect-free, few faulty samples) by encoding the appearance of defect-free samples in a low-dimensional latent vector. The latent vector can reconstruct the original image and can thus capture the distribution of good (i.e., normal) appearances. Any deviation from this distribution is used for the detection of anomalies. In this paper, we demonstrate the classification performance of both methods on the problem of anomalies in the injection moulding process of the routing discs. Although the GANomaly model learns its core model only from defect-free samples and uses faulty samples only for calibration of the detection, we show that on the routing discs domain this can lead to results comparable to the fully supervised method.

The remainder of the paper is structured as follows. First, we present the two evaluated methods in Section 2. A detailed description of the problem domain is presented in Section 3. Comparison of the two approaches is provided in Section 4 with a discussion and conclusion in Section 5.



(a) Acquisition process (b) Data processing / good sample

(c) Bad sample (d) Ground truth mask

Figure 2: (a) The experimental setup for the dataset image acquisition process. (b) The camera view of the injection moulded routing disc with the rotated bounding box in cyan color and final region of interest for defect detection in the dashed green bounding box. (c) Example of a bad sample and (d) the accompanying ground truth mask of faulty parts.

2 Defect-detection approaches

In this section, we briefly describe two state-of-the-art methods for anomaly detection. First, we give a short description of the discriminative method DecisionNet [5], and then we present generative method GANomaly [1].

2.1 DecisionNet

The approach proposed by Tabernik et al. [5] consists of the first stage for segmentation of surface anomalies, and the second stage for the final classification of the image in one of two classes (defect-free or faulty). The trained network achieved perfect results on the DAGM dataset and state-of-the-art results with an average precision of 99.9% on KolektorSDD[1] dataset.

This architecture can be seen in Fig. 1a. The segmentation part of the network is composed of 11 convolutional layers with batch normalization and ReLU activation functions and three max-pooling layers. The final segmentation map is achieved by reducing the previous layer of the network with 1×1 convolutional layer. The classification score is computed in the second stage. The decision network concatenates the last two layers of the segmentation network and puts them through max-pooling and convolutional layers. The final score is computed from the global maximum and average pooling layers which are computed from the segmentation map and previous layer of the decision network. The 66 output neurons are finally combined with linear weights into the final output neuron.

2.2 GANomaly

To address the problem of highly imbalanced datasets, we have used a weakly-supervised approach with a state-of-the-art generative adversarial network for anomaly detection [1]. GANomaly is a fully convolutional generative adversarial network with an encoder-decoder-encoder sub-networks in the generator part. This enables the model to map the input image to a lower-dimensional latent space, which is used to reconstruct the generated output image. Minimizing the reconstruction error metric in the image space and the latent space during the training helps the model to learn the data distribution for the normal samples. This model is learned only on defect-free samples, without any pixel-wise labels. This leads to a larger distance metric from the learned data distribution at inference time, which is indicative of an outlier from the normal distribution and thus produces a larger anomaly score. The abstraction of the GANomaly architecture can be seen in Fig. 1b.

The hypothesis for the model is that when an abnormal image is processed at the inference time, the decoding part of the generator is not able to reconstruct the abnormalities. This stems from the fact that the network was trained by observing only normal (defect-free) samples and its parametrization is not suitable for generating abnormal ones [1]. A large reconstruction error is therefore a good indication of a faulty region.

3 Routing disc dataset

In this section, we present the domain for quality inspection, in particular, the inspection of injection moulded routing disc (see Fig. 2). Injection moulding is a widely used manufacturing process for producing parts with the

[1] http://www.vicos.si/Downloads/KolektorSDD

Table 1: In the first part we are showing the results achieved on the partial images of the whole instances, and in the second part of the table are the aggregated results. TP, TN, FP and FN values are reported at a threshold value where the best F-measure is achieved.

	Partial-view images						*Physical instances*			
Method	mAP	TP	TN	FP	FN	F-measure	TP	TN	FP	FN
DecisionNet	0.998	801	13,428	2	4	0.9963	29	111	1	0
GANomaly	0.993	746	13,433	7	41	0.9695	29	111	1	0

process of injecting molten material into a specially designed mould. Since the parts to be injection moulded must be meticulously designed to facilitate the complexity of the moulding process, a lot of different types of defects can occur[2].

The data acquisition process of the parts was carried out with the Basler aca1600-60gm monochrome camera, a C-mount lens with the focal length of 25 mm, and a high-powered green LED dome illumination as depicted in Fig. 2a. During the acquisition process, the pieces were fixed on a dedicated bed to prevent unwanted movement.

The dataset consists of 112 good samples and 29 bad samples, each captured in a high-resolution image (1602×1202 pixels), which was additionally post-processed to create the final dataset (see Fig. 2b). Each bad sample was annotated with a detailed pixel-wise annotation mask (see Fig. 2d). First, we selected a part of the image with an oriented bounding-box which was incrementally rotated with a step of 3 degrees around the center of the object (depicted with the red arrow in the Fig. 2b). At each step, a smaller area of the image was cut out to provide the final squared area of the part depicted with the dashed green square.

This procedure gave us an intermediate dataset with 13440 partial views of the 112 good samples and 3480 partial views of the 29 bad samples. Image size for the DecisionNet experiment was set to 256×256 pixels. For the GANomaly experiment, the images were reduced to the resolution of 64×64 pixels to make the training times feasible.

Once the samples where cut into parts, only the bad parts with the cutout of the mask containing at least 5% of the total surface area of each partial view were retained. This way we omitted the parts of the bad samples which were sound. The final dataset was comprised of 13440 partial views of the 112 good samples and 803 partial views of the 29 bad samples. A couple of sample images are depicted in Fig. 3.

4 Evaluation

4.1 Evaluation setup

In this section, we present the evaluation of the two reviewed approaches. The training data was prepared as described in the previous section. During the evaluation process 3-fold cross-validation was performed, where all the partial views from a particular physical instance were either in the training set or in the test set.

DecisionNet training is performed for 100 epochs, where on every even step of the learning process the network is shown a faulty sample and a defect-free one on every odd step. In each epoch, the network sees all the faulty images and only a subset of the defect-free ones. During the training of the segmentation sub-network, the fine annotated ground truth mask is provided to guide the training process. Both sub-networks used the cross-entropy loss as in [5].

GANomaly training uses the same training objective as in [1]. The total loss of the model is defined as a weighted sum of the partial losses, *Adversarial*, *Contextual*, and *Latent*, with the model parameters set as in [1]: $\lambda_{adv} = 1$, $\lambda_{con} = 50$, $\lambda_{enc} = 1$, and $|z| = 100$. During the training process, the method uses only defect-free samples. To compute the final anomaly detection score the feature scaling is applied, which is based upon the entire train set including the faulty samples.

4.2 Evaluation results

To evaluate the two different approaches several different classification metrics were measured: (a) mean average precision (mAP), (b) a number of true positive (TP), true negative (TN), false positive (FP) and false negative (FN) samples, and F-measure. In our case, a positive sample represents an image with a present defect, and a negative sample is an image of a defect-free part. The number of TP, TN, FP, and FN samples is reported at a threshold value where the best F-measure is achieved.

Finally, we evaluated the performance of the two competing models according to the classification of complete physical instances. An instance, in this case, is an image of the whole physical piece before it was cut up to smaller squared images as described in Section 3. We pronounced an instance a positive one if at least one partial view was positive, and demanded that all of the partial views of an instance are negative if it is to be a negative instance.

The results are reported in Table 1. In a direct comparison of the discriminative and the generative approach, the DecisionNet achieved mAP of 99.8% where on the other hand GANomaly achieved a slightly lower mAP of 99.3%. A closer look at the FN values reveals that DecisionNet outperformed GANomaly by 4 to 41 miss-classifications. However, once we aggregate the results of the partial views into physical-instance-based metrics, both methods perform comparably well with only a single FP

[2]European Union's Horizon 2020 research and innovation program under grant agreement No. 825030, Project Qu4lity, Kolektor Pilot

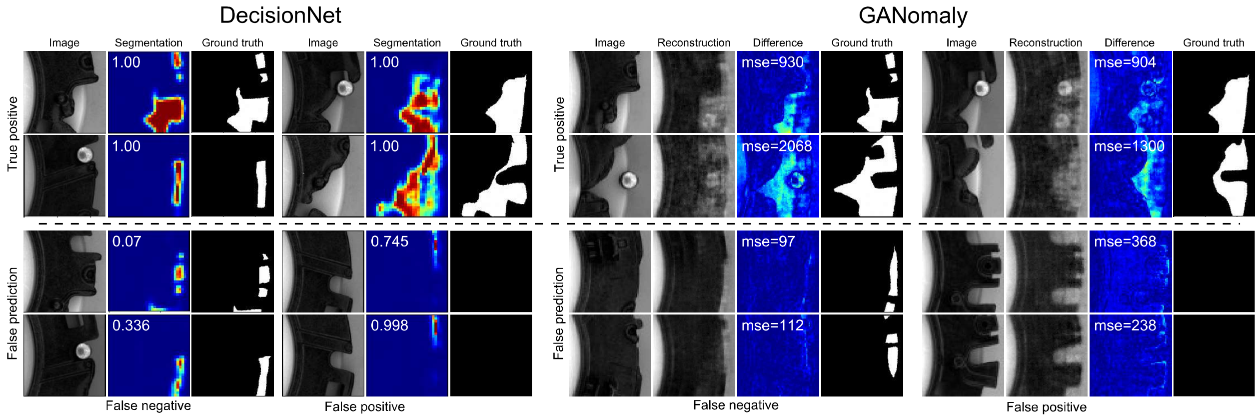


Figure 3: Examples of correct defect detections for DecisionNet (left) and GANomaly (right) shown in the upper two rows, while false negatives and false positives are shown in the lower two rows.

instance. Both methods make a single mistake, though each on a different real-world sample.

Figure 3 depicts several examples of detections for both methods. It also demonstrates the fundamental difference between both approaches. Next to the input image and the ground truth, we depict the segmentation mask produced by the DecisionNet, which serves as a base for the final classification. For GANomaly, we instead depict the reconstructed image and the difference map between the original and reconstructed image, which is, in this case, used to bring the final decision. Since the top examples are not reconstructed well, this indicates the presence of an anomaly in the image.

5 Conclusion

In this paper, we have compared two state-of-the-art methods for anomaly detection on a novel dataset from a real-world industrial domain of injection moulding part. We compared a fully supervised approach, the DecisionNet [5], that uses a supervised discriminative approach to directly learn the appearance of the anomaly, against the generative approach, the GANomaly [1], that uses the generative adversarial network to learn the appearance of defect-free samples. We showed that the discriminative approach out-performed the generative one when measuring on individual partial-view images. However, the GANomaly model performed comparably well when measuring on physical instances. As GANomaly achieves this result by learning the main model only from defect-free samples, this shows a potential for practical use of generative deep-learning models in industrial problems where the acquisition of faulty samples is limited. Moreover, our preliminary analysis also showed that a more intelligent merging of detections from multiple overlapping partial-view images has the potential to further reduce the number of false positive samples.

Our future work will focus on reducing false positive rate using a more advanced fusion of partial results, as well as on exploring different GAN architectures, such as bi-directional GANs [8], and their relevance to the problem of anomaly detection. Furthermore, we will consider addressing the limitation of the image size in existing state-of-the-art GAN-based anomaly detection methods, which has proven to be severely limiting the resolution of the training samples.

Acknowledgements: This work was supported in part by the following research programs: GOSTOP program C3330-16-529000 co-financed by the Republic of Slovenia and the ERDF, ARRS research project J2-9433 (DIVID), and ARRS research programme P2-0214.

References

[1] Akcay, S., Atapour-Abarghouei, A., Breckon, T.P.: Ganomaly: Semi-supervised anomaly detection via adversarial training. In: ACCV 2018. Lecture Notes in Computer Science, Springer (2018)

[2] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. In: Ghahramani, Z., Welling, M., Cortes, C., Lawrence, N.D., Weinberger, K.Q. (eds.) Advances in Neural Information Processing Systems 27, pp. 2672–2680. Curran Associates, Inc. (2014)

[3] Rački, D., Tomaževič, D., Skočaj, D.: A compact convolutional neural network for textured surface anomaly detection. In: WACV 2018. pp. 1331–1339 (2018)

[4] Schlegl, T., Seeböck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised anomaly detection with generative adversarial networks to guide marker discovery. In: Information Processing in Medical Imaging - 25th International Conference, IPMI 2017, Boone, NC, USA, June 25-30, 2017, Proceedings. pp. 146–157 (2017)

[5] Tabernik, D., Šela, S., Skvarč, J., Skočaj, D.: Segmentation-based deep-learning approach for surface-defect detection. Journal of Intelligent Manufacturing (May 2019)

[6] Thoben, K.D., Wiesner, S., Wuest, T.: "Industrie 4.0" and smart manufacturing – a review of research issues and application examples. International Journal of Automation Technology **11**(1), 4–16 (2017)

[7] Weimer, D., Scholz-Reiter, B., Shpitalni, M.: Design of deep convolutional neural network architectures for automated feature extraction in industrial inspection. CIRP Annals - Manufacturing Technology **65**(1), 417–420 (2016)

[8] Zenati, H., Romain, M., Foo, C.S., Lecouat, B., Chandrasekhar, V.R.: Adversarially learned anomaly detection. 2018 IEEE International Conference on Data Mining (ICDM) pp. 727–736 (2018)