ACTA IMEKO
ISSN: 2221-870X
December 2023, Volume 12, Number 4, 1 - 5

Data augmentation for solving industrial recognition tasks with underrepresented defect classes

Lennard Wunsch[1], Katharina Anding[1], Galina Polte[1], Kun Liu[1], Gunther Notni[1]

[1] *Ilmenau University of Technology, Faculty of Mechanical Engineering, Group of Quality Assurance and Industrial Image Processing, Gustav-Kirchhoff-Platz 2, 98693 Ilmenau, Germany*

ABSTRACT

This paper discusses neural network-based data augmentation to increase the performance of neural networks in classification of datasets with underrepresented defect classes. The performance of deep neural networks suffers from an inhomogeneous class distribution in recognition tasks. In particular, applications of deep neural networks to solve quality assurance tasks in industrial production suffer from such unbalanced class distributions. In order to train deep learning networks, a large amount of data is needed to avoid overfitting and to give the network a good generalisation ability. Therefore, a large amount of defect class objects is needed. However, when it comes to producing defect classes, obtaining a dataset for training can be costly. To reduce this costs, artificial intelligence in the form of Generative Adversarial Networks (GANs) can be used to generate images without producing real objects of defect classes. This allows a cost-effective solution for any kind of underrepresented classes. However, the focus of this work is on defect classes. In this paper a comparison of GANs for data augmentation with classical data augmentation methods for simulating images of defect classes in an industrial context is presented. The results show the positive effect of both, classical and GAN-based data augmentation. By applying both methods parallel the best results for defect-class recognition tasks of datasets with underrepresented classes can be achieved.

Section: RESEARCH PAPER

Keywords: Data Augmentation; GAN; Recognition

Citation: Lennard Wunsch, Katharina Anding, Galina Polte, Kun Liu, Gunther Notni, Data augmentation for solving industrial recognition tasks with underrepresented defect classes, Acta IMEKO, vol. 12, no. 4, article 41, December 2023, identifier: IMEKO-ACTA-12 (2023)-04-41

Section Editor: Eric Benoit, Université Savoie Mont Blanc, France

Received July 6, 2022; **In final form** December 12, 2023; **Published** December 2023

Corresponding author: Lennard Wunsch, e-mail: lennard.wunsch@tu-ilmenau.de

INTRODUCTION

Since AlexNet [1] won the ImageNet challenge [2] in 2012, neural networks outperformed classical recognition algorithms in many tasks. With the increase of computational power and easy excess to large amounts of data, it was simple to apply neural networks in recognition tasks. While the depth of networks increases, the architectural design results in fewer parameters within the network which need to be trained. However, there is a need for datasets with a large amount of data to ensure a level of generalization. However, this also means that neural networks are hard to apply to tasks with small datasets or datasets with underrepresented classes and perform rather badly.

Especially in industrial quality assurance applications the total number of objects but also the number of objects within defect classes is usually small, which makes the application of neural networks difficult. Moreover, creating defect objects on purpose is costly and requires a high effort to acquire such data. Therefore, manufacturers want to avoid such measures. Figure 1 displays the quantity distribution of such datasets. While the non-defect class contains a large number of objects compared to the complete dataset, the defect classes only contain a small number of objects.

To address such issues, different approaches are available such as data augmentation to increase the datasets size and reduce the interclass inequality or Transfer Learning to reduce the number of trained parameters. While Transfer Learning [3] was introduced to be able to solve similar tasks with the same network plus training of the fully connected classification layers, data augmentation was used to artificially increase the number of objects present by simulating the variation objects in a virtual space with simple algorithms. This increase of the total number of objects within a dataset reduces the risk of overfitting while training a neural network classifier in addition to reducing the cost to create a homogeneous dataset [4].

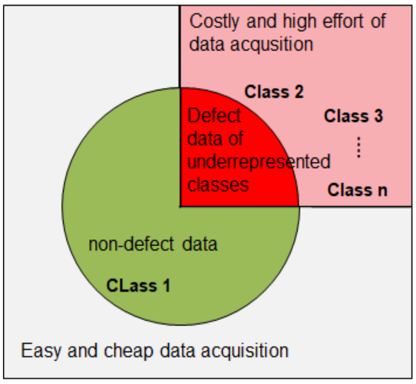

Figure 1. Visualization of the data quantity distribution of a dataset with underrepresented defect classes.

Data augmentation for image processing includes a number of different methods which can be applied in an offline or online manner. Applying methods offline describes methods which are applied to the dataset before training the neural network, while applying methods online, these are applied while training the network to small batches of data. This approach is mostly used in tasks with large datasets to reduce the needed computational power.

Classical methods of data augmentation are for example cropping, mirroring, scaling, random erasing [5], rotation or translational movement. [6] However, with the increased usage of neural networks, GANs [7] were introduced. These network architectures learn to simulate images based on a given dataset.

To evaluate different data augmentation methods, a neural network for defect detection is used and trained based on augmented data in addition to real data. The evaluation process is explained in chapter 4 in detail.

The following paper is structured as follows: first, a closer look at data augmentation methods is given before the datasets of industrial recognition tasks used are presented and the different experiments are described, the results are shown, and a conclusion is given at the end.

2. AN OVERVIEW ON DATA AUGMENTATION METHODS

Data augmentation became an important tool for recognition tasks since it enables the creation of a more diverse and extended dataset. However, it has to be used with care and logically since it can also create object variations which are unnatural and worsen the performance of trained recognition models. Therefore, in addition a certain a priori knowledge of the task at hand, a certain knowledge of the different data augmentation methods and their mathematical background is needed to successfully apply such methods. In [6] and [8] the authors take a closer look into different data augmentation methods.

2.1 Classical methods

Classical methods use simple mathematical operations to vary the image. Cropping a part of the image as well as random erasing and scaling strengthens the model which is more robust since it learns to classify with a less features. However, it can also cause the models performance to worsen if important significant features are missing. This can also give an idea about significant features and redundant features within the dataset [3].

Mirroring or a rotational translation are difficult to apply if convolutional neural networks (CNNs) are used for recognition tasks. Since CNNs use convolution to process data and extract features. Convolution, however, is a translational method and therefore adding rotation to the objects on images creates new filter within the model which can cause the model performance to worsen.

The insertion of additional defects can also be used for data augmentation. By doing so the number of objects within underrepresented defect classes can be increased. This is quite simple in industrial applications when a main defect is a scratch or additional material.

2.2 Generative Adversarial Networks

GANs describes a supervised learning algorithm, introduced into data augmentation by [7]. The algorithm consists of a pair of networks called Generator and Discriminator, which work against each other while training. The generator creates an image based on a random noise pattern while the Discriminator is given either the fake images created by the Generator or a real image from the dataset and decides whether the image given is real or fake. This is given as feedback to the Generator Therefore, while the Generator learns to create more realistic images, the Discriminator becomes better in distinguishing between fake and real. One disadvantage of GANs is that a large amount of computing power and time is required to generate realistic data. Training GANs can take a large amount of time before created images can be of use.

In this work, StyleGAN developed by [9] was used for simulated data generation. Compared to traditional GANs, StyleGAN uses controllable vectors in addition to a randomly generated input to manipulate the features of generated images. Moreover, a mapping-network consisting of fully connected layers is used to create the controllable input vector. A closer overview of StyleGAN including a comparison with traditional GANs is given in [9].

3. DATASETS

In this work two different datasets with industrial background are used to compare different data augmentation methods. Both are classical defect class recognition tasks from quality assurance in industry, where an object needs to be inspected after production. Therefore, each dataset contains one non-defect class. Dataset one contains macro plastic injection moulded objects of LED housings for automotive applications. The defect classes can be separated into a class of objects where the nozzle was set up too hot, objects with an incorrect granulate composition and objects where the holding pressure was set up too small.

Dataset two contains automatic visual inspection of metallic surfaces produced by milling. It in Addition to the non-defect class, a class of images which show scatter marks and one which show longitudinal rills are also part of this dataset.

Figure 2 displays examples of the different defect classes as well as one example of a perfectly produced object. While some of the differences between some defect classes and the perfect class are easily recognized, such as the scatter marks on the milled rings or the incorrect granulate composition at the LED housing, others are more difficult to recognize such as the LED Housings where the holding pressure was set up to small.

Both datasets contain images with only single type defects and are labelled as such. In both datasets localization and severeness of defects are of less importance, since one can conclude the source independent from the defect localization or severeness. Moreover, correction measures are be introduced solely based on the type of defect as well.

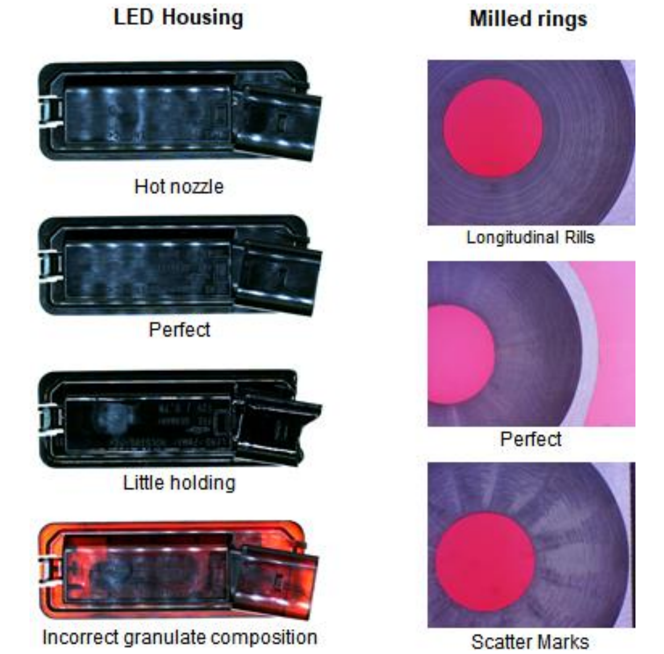

Figure 2. Objects of each defect class and the perfect class of both datasets.

4. EXPERIMENTS

Different experiments were conducted with different classical data augmentation methods and the StyleGAN. To be able to compare the results, the fake images were given the same network. The VGG19 [10] network was chosen as a base for defect detection, due to its simple architecture and strong performances. VGG19 is an established network, which contains 19 feature-processing layers with pooling layers in between. Transfer Learning was applied on the defect detection network by initializing frozen pretrained weighs based on ImageNet Dataset. By convolution, 2D-features are extracted, and two fully connected layers are used for fine tuning, before the output layer is used for classification. This results in approximately 26 million trainable parameters. Overfitting was avoided by introducing dropouts and a reduced training period. Training the defect detection was stopped after 20 epochs when recognition rates started to saturate.

The datasets were divided into a train set (60 %), a validation set (20 %) and a test set (20 %). While the total number of objects within the test set was kept constant the training set and validation set were adjusted to the total number of objects. For validation, k-fold validation was used. As benchmark (experiment No. 1) the model was given only real images of the datasets to be able to ratify whether data augmentation increases or decreases the model's performance.

Different subsets of the datasets were created using only classical methods of data augmentation (No. 2) and adding to original real data, adding only the StyleGAN (No. 3) produced data and adding a mixture of both types of data augmentation produced data (No. 6). To be able to compare the methods, the number of used images was consistent. Additional experiments were conducted with twice the number of images using the classical methods (No. 4) and StyleGAN (No. 5). In addition, an experiment was conducted, again using a mixture of twice as many classically created images and GAN-created images, to investigate the influence of a large increase in artificially created images (No. 7). Each subset is part of one experiment and shall give an idea about the influences of different data augmentation methods on recognition performances of a neural network.

Moreover, since GAN training can be conducted using an untrained raw set of networks or Transfer Learning [3] can be applied, investigations concerning the image quality of the created images were done as well. This investigation was done before conducting other experiments. For Transfer Learning the creation of one class was trained before the information was used for training the creation of other classes. This way the training process can be accelerated.

To reduce the GAN training time the image sizes have been reduced to 256 x 205 pixel in the milled rings dataset and 512 x 183 pixel in the LED housing dataset. To be able to train GANs, a large amount of computational power is needed. Therefore, the training was conducted using a NVIDIA GeForce RTX 2080 Ti graphics card. GAN training for dataset one containing LED housings stopped at 15 10^6 images, while GAN training for dataset two containing metallic surfaces was stopped at 25 10^6 images. Both stops were determined empirically.

Table 1 shows the data augmentation methods used in the corresponding experiments. These numbers will be used to allocate the results of each experiment. The total number of images used in each experiment for the corresponding dataset is also given. The balance between classes was kept constant, so that the number of total images was increased but the percentage of each class contributing to the dataset is constant. With increased use of data augmentation methods, the number of images also increases. To ensure a robust training process, for each experiment the defect detection model was initialized and trained five times to. For comparison mean recognition rate and standard deviation were calculated based on all five runs for each experiment.

5. RESULTS

The Benchmark was set to be able to see whether the used data augmentation methods enable a well-known network architecture to increase its performance. First examples of GAN-generated images are shown to get an impression of the quality of the images generated with GAN. Then the recognition rate and standard deviation for each experiment performed are presented to compare the results. An overview of the training times needed is given at the end before the conclusion.

By comparing images created by non-pretrained GANs to pretrained GANs and GAN training times in Table 4 (longer training times for “Longitudinal Rings” and “Little holding”

Table 1. Overview of the conducted experiments.

Experiment No.	Data augmentation methods	Total number of images LED housing/Milled rings
1	Benchmark	680 / 275
2	Real images and classical methods	1224 / 491
3	Real images and GAN created images	1224 / 491
4	Real images and double the number of classical methods	1768 / 711
5	Real images and double the number of GAN created images	1768 / 711
6	Real images, classical methods and GAN created images	1768 / 711
7	Real images and each double the number of images of classical method and GAN created images	2856 / 1145

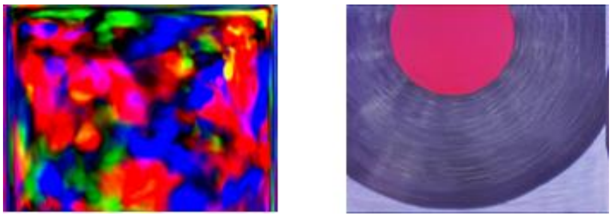

Figure 3. Start of training without (left) and with transfer learning (right) of the metal data set.

compared to the other classes when using transfer learning), the need for transfer learning became clear. Figure 3 shows the difference in start of training GANs with and without transfer learning. [Image created by using a raw set of GANs (left) and an image created using pretrained GANs via transfer learning (right)]. While the right image shows a realistic image of milled rings, the left displays a noisy pattern as start of training. Thus, in the experiments conducted (see Table 1), the first data-augmented class ("Longitudinal Rings" in the metal data set of milled rings and "Little holding" in the LED housing data set) was performed without transfer learning and all other classes of the two data sets were performed using transfer learning.

Figure 4 shows two images of the data-augmented milled-rings dataset. Both were artificially generated by StyleGAN and contain longitudinal rills. Comparing the images with the example of milled rings with longitudinal rills in Figure 2, both seem to be authentic and have similar characteristic.

Figure 5 shows artificially generated images of LED housings. The images can also not be differentiated from images of real objects.

Mean Recognition rate and Standard deviation of the evaluation process using VGG 19 are given in Table 2 and Table 3 for each dataset respectively. While Table 2 presents the results of the experiments conducted on the LED Housing dataset, Table 3 shows the results of the experiments conducted on the milled rings dataset. When evaluating the results, the reader must always consider the increase of images with the addition of data augmentation. The reader must keep in mind that the benchmark (experiment No. 1) was trained using the smallest number of images of all experiments.

Table 2 shows in general, that each data augmentation method increased the model performance and made it also more robust. Moreover, it can be seen, that by adding GAN created images the model is able to slightly outperform the one where classically created images were added. By adding more artificially created images, the model's performance increases even further. However, the recognition rates of the experiments with a doubled number of artificial images are too close to each other to conclude one method superior to the other. Also adding even more artificial images does not have a big effect on the recognition rate.

Figure 4. Examples of GAN-generated images of milled rings.

Figure 5. Style-GAN generated LED housing images showing top down: hot nozzle effects, perfect, little holding effects and effects of incorrect granulate composition.

Table 3 displays the results of the experiments conducted on the milled rings dataset. The experiments on the milled rings dataset also show an increase on the network's performance with the addition of data augmentation.

Compared to the performance on the LED housing dataset, the performance on the milled rings dataset is improving strongly by up to 8.36 %.

However, in this application the classical methods slightly outperform the GAN-based approach. Moreover, the best and most robust performance was reached when the number of objects was increased even further. This can be either a result based on simply a larger number of objects or also due to the bigger diversity within the datasets classes.

In total training the GAN for creation of milled rings took 45 days, 8 hours and 47 minutes while training for the LED housing image creation took 68 days, 14 hours and 41 minutes. Table 4 given an overview of the time used per class. The effect of transfer learning can be seen when comparing either the time for

Table 2. Mean Recognition rates and standard deviation of experiments No. 1-7 conducted on the LED housing dataset.

Experiment No.	Mean Recognition rate %	Standard deviation %
1	92.79	3.69
2	93.09	1.43
3	94.12	1.16
4	95.10	3.01
5	95.15	1.69
6	95.00	1.90
7	95.29	1.23

Table 3. Mean Recognition rates and standard deviation of experiments No. 1-7 conducted on the Milled rings dataset.

Experiment No.	Mean Recognition rate %	Standard deviation %
1	82.91	5.24
2	86.55	2.44
3	86.18	4.74
4	90.55	1.99
5	89.09	2.23
6	89.09	1.82
7	91.27	1.52

Table 4. Overview of the GAN training times per dataset and class.

Dataset	Class	Training time		
		days	hours	minutes
Milled rings	Longitudinal Rings	18	22	35
	Perfect	13	23	21
	Scatter marks	12	20	51
LED housing	Hot nozzle	10	18	44
	Perfect	11	3	25
	Little holding	35	20	38
	Incorrect granulate composition	10	19	54

the "Longitudinal Rings" class or the "Little holding" class to the times of training other classes.

6. CONCLUSION

This paper presented a comparison between classical data augmentation methods and the new neural network-based data augmentation by GAN applied on industrial data. It shows that both methods increase the performance of neural network-based recognition and also increases the networks robustness. The improvement on both different applications shows a trend, which can be transferred to other applications in industry as well.

Moreover, data augmentation not only improves a networks performance, but also increases its robustness. Therefore, the general use of data augmentation can strongly improve a networks performance on industrial applications for quality assurance on manufactured objects.

While in one application the classical methods outperform the GAN based approach, the best results were reached when combining both approaches for increase the datasets size and variety of the classes. This can be explained by the different advantages and disadvantages of both methods. Future research needs to be done to precisely determine the reason of the better performance of the combination of classical and GAN-based approaches. Furthermore, to rate the vicinity of reality of simulated images, such as GAN generated images, further investigations must be done with experiments containing the same number of images but a different ration of real and simulated images. Moreover, the process of overfitting needs to be investigated in future works as well. By monitoring the training process of StyleGAN and evaluation at different stages, the effect of overfitting on the quality of GAN-based simulated data should be further investigated.

ACKNOWLEDGEMENT

The research project OptoCheck, which forms the basis of the investigations of plastic injection moulded LED housings, was funded by the Free State of Thuringia and co-financed by the European Union under the European Regional Development Fund (ERDF). The investigations of macro plastic injection moulded components (led housing) are based on the project "OptoCheck – Novel method for inline inspection of dimensional accuracy and surface quality on complex components in mechanical engineering and automotive".

The samples of metal surfaces with different surface defects, like scratches and chatter marks were produced by our project partner the society for production engineering and development (GFE e.V.) in Schmalkalden, Germany. Our special thanks go to Dr. Daniel Garten and his colleagues for the production of metal surfaces and the image acquisition for building the dataset of metal surfaces with and without defects.

The specific investigations of data augmentation with GAN were carried out in the project ThurAI. This project is also funded by the Free State of Thuringia and co-financed by the European Union under the European Regional Development Fund (ERDF).

We thank the funding authorities for the financial support of this work.

The responsibility for the research content lies with the authors.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, G. E. Hinton, ImageNet classification with deep convolutional neural networks, in Proc. of the 25th Int. Conf. on Neural Information Processing Systems, vol. 1, Lake Tahoe, Nevada, USA, 3-6 December 2012, pp. 1097–1105.

[2] O. Russakovsky, J. Deng, H. Su, (+8 more authors), ImageNet large scale visual recognition challenge, Int. Journal of Computer Vision, vol. 115 (2015) 3, pp. 211-252. DOI: 10.1007/s11263-015-0816-y

[3] S. J. Pan, Q. Yang, A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering, vol. 22 (2010) 10, pp. 1345-1359. DOI: 10.1109/TKDE.2009.191

[4] M. Cogswell, F. Ahmed, R. B. Grishick, C. L. Zitnick, D. Batra, Reducing overfitting in deep networks by decorrelation representations, Int. Conf. on Learning Representation, 19 November 2015, arXiv 2016. DOI: 10.48550/arXiv.1511.06068

[5] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, Proc. of the AAAI Conf. on Artificial Intelligence, vol. 34 (2020) 07, p. 13001-13008. DOI: 10.48550/arXiv.1708.04896

[6] S. Yang, W. Xiao, M. Zhang, S. Gou, J. Zhao, F. Shen, Image data augmentation for deep learning: A survey, arXiv: 2204.08610 DOI: 10.48550/arXiv.2204.08610

[7] I. J. Goodfellow, Jean Pouget-Abadie, M. Mirza, (+5 more authors), Generative adversarial nets, in Proc. of the 27th Int. Conf. on Neural Information Processing Systems – vol. 2, MIT Press: Montreal, Canada, 2014, pp. 2672–2680.

[8] C. Shorten, T. M. Khoshgoftaar, A survey on image data augmentation for deep learning, Journal of Big Data, vol. 6 (2019) 1, p. 60. DOI: 10.1186/s40537-019-0197-0

[9] T. Karras, S. Laine, T. Aila, Supplemental material: A style-based generator architecture for generative adversarial networks, 2019.

[10] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv: 1409.1556, 2014. DOI: 10.48550/arXiv.1409.1556