Automated Dataset Enhancement Using GAN for Assessment of Degree of Degradation around Scribe

1st Petr Dolezel*, 2nd Veronika Rozsivalova†, 3rd Marek Pakosta‡, 4th Dominik Stursa§
Faculty of Electrical Engineering and Informatics
University of Pardubice
Pardubice, Czech Republic
Email: *petr.dolezel@upce.cz, †veronika.rozsivalova@student.upce.cz,
‡marek.pakosta@student.upce.cz, §dominik.stursa@upce.cz

***Abstract*—Coil coating is a method of applying an organic coating material to a rolled metal strip substrate in a continuous automated process. It is used to provide a high quality, durable finish to a variety of surfaces. The degradation resistance of coil-coated materials is assessed according to European Standard EN 13523-8 by exposing a coil-coated test specimen to a salt fog at a defined temperature for a defined period of time. After this process, a sample is tested according to the International Organisation for Standardisation ISO 4628 standard to determine the degree of degradation. In this study, a GAN-based technique for automated training set enhancement is proposed to assess the degree of degradation around a scribe. The presented technique is capable of enhancing a manually generated dataset of images with synthetic samples to help refine the performance of the area degradation detector.**

***Index Terms*—coil coating, delamination, degradation, semantic segmentation, deep learning, generative adversarial network**

I. Introduction

Coil coating is a method employed in a continuous automated process to apply an organic coating material onto a rolled metal strip substrate. Its purpose is to achieve a uniform, high-quality, and long-lasting finish on metal surfaces [1], [2]. Coil coating serves as a crucial solution for applications requiring a durable and enduring finish, such as building exteriors, metal roofs, wall panels, garage doors, office furniture, vending machines, food service equipment, and others. Additionally, it finds utility in advanced and intricate coatings like cool metal roofing materials, smog-eating building panels, antimicrobial products, anti-corrosive metal parts, and solar panels.

Therefore, coil coating offers a thin but durable and flexible protective layer that effectively safeguards the material against corrosion. However, this protective layer is susceptible to mechanical damage, such as scribes or scratches, which can result in irreversible changes to the material due to exposure to water, sunlight, salt, or corrosive gases [3]. Such damage can manifest in various forms, ranging from chalking and blistering to flaking or rusting of the coated material. Consequently, it becomes essential to evaluate the performance of coated surfaces under conditions that accurately simulate outdoor exposure.

The assessment of the degradation resistance of coil coated materials adheres to the European Standard EN 13523-8 titled ”Coil coated metals. Test methods. Resistance to salt spray (fog).” This standard involves subjecting a test specimen treated with coil coating to a salt fog at a predefined temperature and duration. Following this process, the specimen undergoes testing according to the International Organization for Standardization (ISO) 4628 standard titled ”Paints and varnishes. Evaluation of degradation of coatings. Designation of the quantity and size of defects and intensity of uniform changes to appearance.” In simpler terms, the objective is to evaluate the extent of degradation on the test specimen, as illustrated in Fig. 1. The degree of degradation is subsequently determined by calculating the ratio of the degraded area to the total area of the specimen.

Typically, the area of degradation is determined manually, as indicated in [3], [4]. Surprisingly few approaches using automatic image processing-based methods have been presented so far. An example of the specific method of image processing for degradation evaluation can be found in works [5], [6]. Other authors proposed using office scanners in combination with commercial software such as Adobe Photoshop [7]. However, the common disadvantage of these approaches is the limitation to a specific type of material, surface, or color of the sample. In contrast, the industrial practice requires methods that are generally applicable to the widest possible range of materials. With this requirement in mind, the authors of this study proposed a method based on semantic segmentation using fully-convolutional neural networks [8]. The results presented in that work clearly showed that U-shaped fully convolutional networks were very capable of automatically detecting the degradation area of surfaces treated with coil coating.

A significant drawback of all data-driven methods is the need for a large and complex training set, well covering

The work has been supported by the SGS grant at the Faculty of Electrical Engineering and Informatics, University of Pardubice, Czech Republic. This support is very gratefully acknowledged.

Test specimen removed from salt fog chamber.

Degradation detection.

Fig. 1: Examples of degraded area determination.

significant states of the problem to be solved. Especially for tasks based on semantic segmentation, it is very challenging not only to measure the dataset but also to annotate it correctly. This study, therefore, looks specifically at the possibility of automatically extending a manually created dataset using the Generative Adversarial Network (GAN). GAN is an unsupervised learning framework used especially for generative tasks. It consists of two networks – the generator and the discriminator – that compete against each other in order to produce better results. The generator network produces data that are similar to the training data, while the discriminator network tries to differentiate between the generated data and the real data. When both networks are trained in parallel, the generator is able to produce realistic data that are increasingly difficult for the discriminator to distinguish. The framework was first presented in 2014 [9] and has been applied to a variety of tasks, such as image classification, object detection, image segmentation, and natural language processing, as well as to generate realistic images, videos, and music.

Hence, the aim of this study is to propose a pilot GAN-based technique for automated training set enhancement for a semantic segmentation task. The overall pipeline of the technique is shown in Fig. 2.

II. Materials and Methods

The aim of this study is to propose a method to enhance the dataset using GAN to assess the degree of degradation around a scribe. Therefore, the dataset created for this task in [8] is used in the experiments. The dataset is then enhanced using GAN, which is designed to generate both inputs and targets related to the images present in the original dataset (see Fig. 1 for some data examples). To show the effect of the dataset enhancement, the original dataset and the enhanced dataset are used to train a semantic segmentation neural network. The U-net [10] is used for semantic segmentation because it is a well-known architecture that has proven its qualities in many real-world applications.

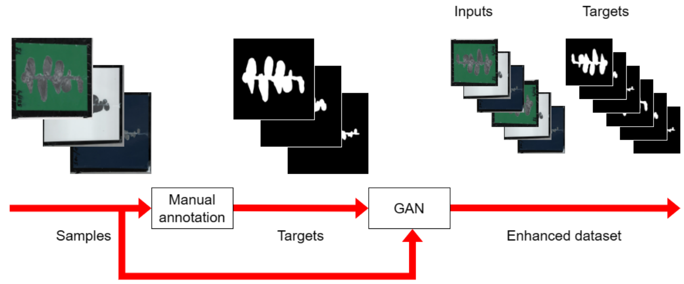


Fig. 2: Automated training set enhancement. Manually created limited training set is multiplied by generative adversarial network.

A. Original dataset

To develop a robust data-driven approach for automated detection of coating degradation, it is imperative to obtain a substantial collection of diverse annotated samples exhibiting variations in color, asperity, and reflectivity. Furthermore, these samples should encompass different levels of degradation. In the original study by Rozsivalova et al. [8], a total of 586 coated samples measuring 150×100 mm were meticulously prepared. The samples featured coatings in various colors, including black, white, green, gray, orange, red, brown, blue, dark blue, and yellow, with both fine and coarse (textured) variations.

In order to expose the uncoated material, a small horizontal scratch, measuring 0.5 mm in width, was intentionally made through the coating of each sample using an iron nail. Subsequently, the samples were subjected to a defined period of exposure in a salt fog chamber, lasting for durations of 120 h, 240 h, 480 h, 720 h, and 1440 h. Following the exposure period, the samples were thoroughly cleaned and scanned using an office scanner. Lastly, each individual sample underwent manual annotation to obtain the target segmentation image required for training the area degradation detector.

B. U-net semantic segmentation neural network

U-Net is a symmetric dense pixelwise prediction architecture designed originally for biomedical image segmentation. It is an encoder-decoder network consisting of a contraction path and an expansion path connected by a series of convolutional layers. U-Net uses skip connections to concatenate feature maps from the contraction and expansion paths, allowing the model to better localize features and capture detailed information from the input image. U-Net has been applied to many types of imaging tasks, with applications ranging from brain tumor segmentation [11] to cell tracking [12].

For the purpose of this study, U-net is modified to work with an $(288 \times 288 \times 3)$px RGB image at the input and provide a $(288 \times 288 \times 1)$px black-and-white image at the output. The architecture is depicted in Fig. 3.

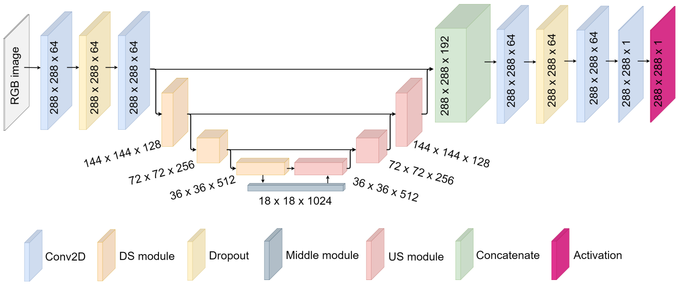


Fig. 3: U-net adapted for the assessment of the degree of degradation around a scribe. See [10], [13] for the detailed description of the architecture.

C. Generative adversarial network for dataset enhancement

The goal of this study is to propose a GAN-based method to enhance the original dataset of input and target images of diverse annotated samples with different colors, asperity, or reflectivity. GANs use two neural networks (generator and discriminator) to generate synthetic data. The generator network takes a random noise vector as an input and produces a synthetic image that should be indistinguishable from a real image. On the other hand, the discriminator network takes both real and synthetic images and attempts to classify the input as either real or fake. The generator and discriminator networks are trained together in a competitive manner, where the generator aims to produce images that the discriminator is unable to distinguish as fake ones, while the discriminator tries to detect fake images that were produced by the generator.

For the purpose of this study, GAN is adapted to work with $(288 \times 288 \times 4)$ data structures. These structures are created as a concatenation of an RGB image representing the input image and a black-and-white image representing the target image. The configuration is illustrated in Fig. 4. The chosen discriminator and generator architecture is shown in Fig. 5 and Fig. 6, respectively. Note that the used architectures are not explicitly specified in order to keep the text compact. Nevertheless, the dimensions of the data streams and the individual layer types are indicated in the figures.

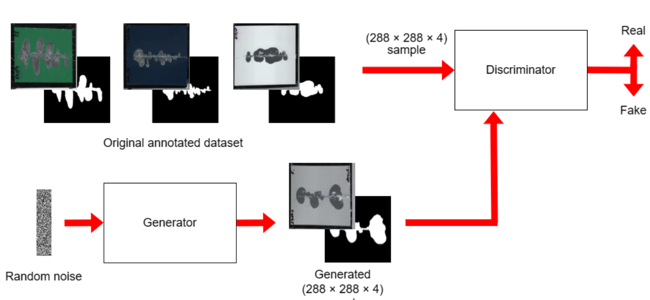


Fig. 4: Generative adversarial network for dataset enhancement.

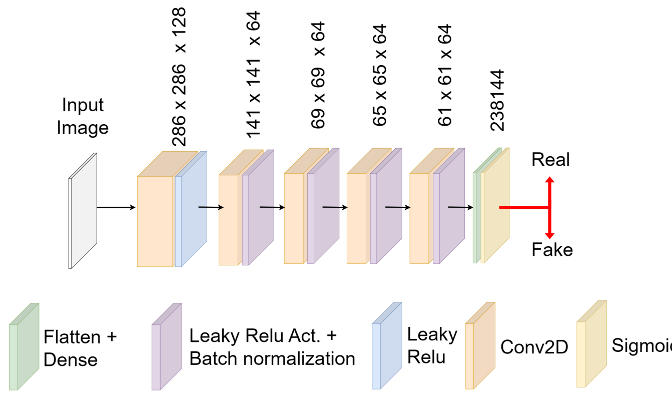


Fig. 5: Discriminator architecture for generative adversarial network.

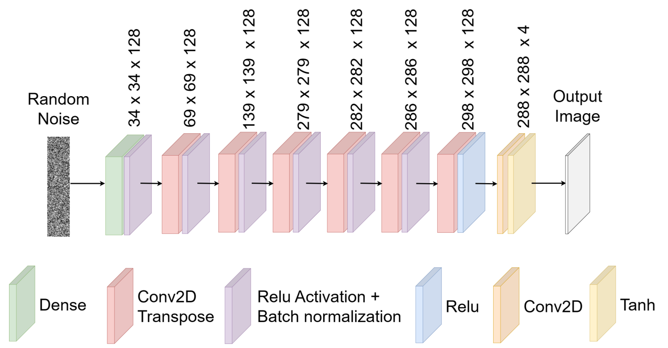


Fig. 6: Generator architecture for generative adversarial network.

D. Evaluation metrics

A comprehensive set of evaluation metrics must be considered to clearly demonstrate the benefits of automated dataset expansion. A first choice for the evaluation of the classification performance is calculation of accuracy over a testing set (a dataset independent of the training set). The task of assessing the degree of degradation is essentially a pixel classification.The task dealt with during the assessment of the degree of degradation around a scribe is basically a pixelwise classification of the image content. For the classification of the degradation area, a true positive pixel is a pixel labelled as white in the target image as well as in the output image. A false positive pixel is labelled as white in the output, but is black in the target image. A true negative pixel is a pixel labelled as black in the target image as well as in the output image. Ultimately, a false negative pixel is labelled as false in the output image, but is white in the target image. Then, the accuracy is given as

$$\text{Accuracy} = \frac{|\text{TP}| + |\text{TN}|}{|\text{TP}| + |\text{FP}| + |\text{TN}| + |\text{FN}|}, \tag{1}$$

where TP, FN, FP and TN is the number of true positive, false negative, false positive and true negative pixels in the tested sample respectively.

To evaluate the classification performance comprehensively, additional metrics are also considered:

$$\text{Precision} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FP}|}, \tag{2}$$

$$\text{Recall} = \frac{|\text{TP}|}{|\text{TP}| + |\text{FN}|}, \tag{3}$$

$$\text{F1-score} = \frac{2}{\text{Recall}^{-1} + \text{Precision}^{-1}}. \tag{4}$$

III. Experiments and results

A. Setup of experiments

All experimental models were implemented using Python 3.9 with TensorFlow 2.0 and Keras framework. Experiments were performed using the following hardware specification: Intel Core i5-8600K (3.6 GHz) processor, 16 GB DDR4 (2666 MHz) internal memory, NVIDIA PNY Quadro P5000 16 GB GDDR5 PCIe 3.0 (2560 CUDA cores) video card, SATA M.2 512 GB SSD.

B. Training using original dataset

Before an enhanced dataset, an original dataset was separately used for training. U-net model was trained using Adam optimizer, as it is generally considered to perform acceptably in most cases [14], [15]. The weights were initialized randomly with Gaussian distribution. Binary cross entropy was used as the loss function. From the original dataset of 586 input-target pairs, 130 pairs were set aside to form the testing set. The rest of the dataset was used for training, with 15 % of it used for validation. The experiments were performed ten times to suppress the stochastic nature of training. The best instances, considering the loss function over the validation set, were then used for further evaluation. All the parameters of the training are summarized in Table I.

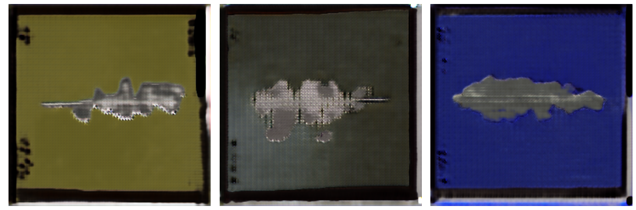
TABLE I: Parameters of the training

Input shape	288 x 288 x 3
Training algorithm	Adam optimizer
Number of training experiments	10
Number of samples	456
Validation split	0.15
Initialization	Normal distrib. (mean = 0, std = 0.05)
Number of epochs	500
Batch size	16
Criterion for resultant model	Loss function value over validation set
Learning rate α	0.001
Exponential decay rate 1 β_1	0.9
Exponential decay rate 2 β_2	0.999

C. Original dataset enhancement

In order to enhance the original dataset, GAN model was prepared according to the pipeline in Fig. 4. The generator–discriminator pair was trained with similar parameters as in the previous case, the only difference being the batch size, which is reduced to 10 for memory complexity reasons.

The generator from the trained GAN architecture was then used to enhance the original dataset. Specifically, 300 synthetic input-target pairs were generated using random input signal. Three examples from the generated data are shown in Fig. 7.



Synthetic input images.

Synthetic target images.

Fig. 7: Examples of generated sythetic data. Each image in second row corresponds to image in first row.

D. Training using enhanced dataset

U-net model was trained using Adam optimizer from scratch using the enhanced dataset. Here, the testing and validation set remained the same as in the previous case, but the training set was enriched with 300 additional synthetic pairs as described above, together forming a 756 samples in total. Aside from enriched dataset, the rest of the parameters were identical to the values given in Table I.

The resultant values of the considered metrics for training with both the original dataset and the enhanced dataset are shown in Table II.

TABLE II: Resulting values of the considered metrics over testing set

Dataset	Accuracy	Precision	Recall	F1-score
Original	0.9893	0.8906	0.8777	0.8841
Enhanced	0.9919	0.9005	0.9066	0.9035

Considering the values in Table II, it is obvious that the enhancement of the training set with synthetic samples clearly increased the detector performance in terms of the area degradation detection. While the increase in the accuracy metric is negligible, the increase in the recall and F1-score metrics is significant. It should be emphasized that the synthetic data were only present in the training set. The same testing set was used for both training experiments.

IV. Conclusions

Here, a GAN-based technique for automated training set enhancement for a semantic segmentation task is proposed. Specifically, the assessment of the degree of degradation around a scribe is dealt with. In this study, it is shown that GAN can be used to enhance the training set for semantic segmentation task when the original dataset does not contain enough data. However, the study should only be seen as a pilot experiment that opens up further directions of research. In the near future, the effect of the amount and ratio of synthetic data in the training set on the quality of the gained detector will need to be investigated. It is also possible to optimize the discriminator and generator in the GAN architecture to achieve the best possible results.

References

[1] A.-S. Jandel, “Sustainability in the coil coatings sector [nachhaltigkeit in der coil-coating-branche],” *JOT, Journal fuer Oberflaechentechnik*, vol. 59, no. 8, pp. 10–13, 2019.

[2] The National Coil Coating Association, “NCCA promotes growth of coil coating industry,” *Light Metal Age*, vol. 78, no. 5, pp. 66–67, 2020.

[3] A. Bastos and A. Simões, “Effect of deep drawing on the performance of coil-coatings assessed by electrochemical techniques,” *Progress in Organic Coatings*, vol. 65, no. 2, pp. 295–303, 2009.

[4] J. Hanus, “Selection and evaluation of singlelayer coating compositions in corrosive environments,” *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, vol. 59, no. 5, pp. 53–64, 2011.

[5] P. Kapsalas, M. Zervakis, and P. Maravelaki-Kalaitzaki, “Evaluation of image segmentation approaches for non-destructive detection and quantification of corrosion damage on stonework,” *Corrosion Science*, vol. 49, no. 12, pp. 4415–4442, 2007.

[6] E. C. Cringasu, A. Dragomirescu, and C. A. Safta, “Image processing approach for estimating the degree of surface degradation by corrosion,” in *2017 International Conference on ENERGY and ENVIRONMENT (CIEM)*, Oct 2017, pp. 275–278.

[7] P. Blanchard, D. Hill, G. Bretz, and R. McCune, *Evaluation of Corrosion Protection Methods for Magnesium Alloys in Automotive Applications*, 2014, vol. 9781118858943.

[8] V. Rozsivalova, P. Dolezel, D. Stursa, and P. Rozsival, “Sequence of u-shaped convolutional networks for assessment of degree of delamination around scribe,” *International Journal of Computational Intelligence Systems*, vol. 15, no. 1, 2022.

[9] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” vol. 3, no. January, 2014, Conference paper, p. 2672 – 2680.

[10] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351, pp. 234–241, 2015.

[11] A. M. Gab Allah, A. M. Sarhan, and N. M. Elshennawy, “Edge u-net: Brain tumor segmentation using mri based on deep u-net model with boundary information,” *Expert Systems with Applications*, vol. 213, 2023.

[12] L.-X. Yuan, H.-M. Xu, Z.-Y. Zhang, X.-W. Liu, J.-X. Li, J.-H. Wang, H.-B. Cui, H.-R. Huang, Y. Zheng, and D. Ma, “High precision tracking analysis of cell position and motion fields using 3d u-net network models,” *Computers in Biology and Medicine*, vol. 154, 2023.

[13] P. Dolezel, P. Skrabanek, D. Stursa, B. Baruque Zanon, H. Cogollos Adrian, and P. Kryda, “Centroid based person detection using pixelwise prediction of the position,” *Journal of Computational Science*, vol. 63, 2022.

[14] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *CoRR*, vol. abs/1412.6980, 2014. [Online]. Available: http://arxiv.org/abs/1412.6980

[15] E. M. Dogo, O. J. Afolabi, N. I. Nwulu, B. Twala, and C. O. Aigbavboa, “A comparative analysis of gradient descent-based optimization algorithms on convolutional neural networks,” in *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, 2018, pp. 92–99.