Available online at www.eccomasproceedia.org
Eccomas Proceedia SMART (2023) 1221-1233

ECCOMAS
Proceedia

X ECCOMAS Thematic Conference on Smart Structures and Materials
SMART 2023
D.A. Saravanos, A. Benjeddou, N. Chrysochoidis and T. Theodosiou (Eds)

INSPECTION OF SURFACE DEFECTS IN METAL PROCESSING INDUSTRY USING UNET-BASED ARCHITECTURES

Lampros Leontaris[1,*], Nikolaos Dimitriou[1], Apostolos Nikolousis[2], Dimitrios Tzovaras[1] and Elpiniki Papageorgiou[3]

[1]Information Technologies Institute
Centre for Research and Technology Hellas
Thessaloniki 57001 Greece

[2] PYRAMIS METALLOURGIA A.E.
17th Km Thessaloniki - Serres
P.O. Box: 10278, Thessaloniki

[3] Dept. of Energy Systems, Faculty of Technology
University of Thessaly, Geopolis Campus
Larissa 41500, Greece

*Corresponding author: Lampros Leontaris
e-mail: lleontar@iti.gr

Abstract. Surface inspection is a critical procedure of quality control during metal processing. The raw material is processed in various production stages including cutting, punching, trimming, scrubbing and polishing. Each of these stages introduce different parameters that are controlled by human operators. Sub-optimal parametrization of the machining process causes the appearance of defects in the metallic surface. Quality control typically involves human inspection at the last stages of the manufacturing where the defects cannot be repaired, while due to surface specularity, manual inspection is error-prone and time consuming. In this work, we propose a vision-based system to automate quality control, that comprises two machine vision sensors, an illumination system and a mounting frame. We investigate segmentation-based Deep Learning approaches using the UNet architecture for the localization of defects in images. We validate the proposed method in a real sink manufacturing case using high resolution image data that has been collected during production, containing various types of surface defects. We examine the effect of different network backbones and loss functions in the performance of the examined approaches, using precision, recall and f-measure metrics. The findings demonstrate the potential of the system to be used in the production in order to provide accurate detections and thus accelerate inspection.

Key words: Deep Learning, Surface Inspection, Sink manufacturing, Quality Control

Peer-review under responsibility of the organizing committee of SMART 2023.
doi: 10.7712/150123.9870.444670

1 INTRODUCTION

With the advent of Deep Learning (DL) and Industry 4.0, automated quality control solutions in Smart Manufacturing have been in the focus of industrial research. Even if Artificial Intelligence (AI) solutions have already reached a level of maturity in several application domains [1], different industrial scenarios introduce challenges risen by the inspected objects' properties and specific production requirements. On one hand, an automated defect detection system must cope with a wide variety of surface defects, with different shapes and sizes, while on the other hand, time restrictions are typically involved, requiring inspection to be completed in a short time. These challenges have made traditional image processing approaches for defect detection, impractical and unreliable. DL have demonstrated efficiency and robustness in a wide range of industrial domains and specifically in surface inspection for weld defects [2], casting defects in auminum alloys [3], road pavement cracks [4], fabric defects [5], lithium-ion battery defects [6], hard metal defects [7], automotive inspection [8, 9] and PCB defects [10].
In our work, we focus on defect detection methods for metallic surfaces in sink industry. During sink manufacturing, surface inspection is the most critical process of quality control, while it is typically performed in the final stages of manufacturing, where recovery actions cannot be efficiently applied. This causes increased scrap rates but even when reworking is feasible, production delay raises manufacturing costs. Initially, the raw material is processed through a multi-stage production line where cutting, punching, trimming, scrubbing and polishing of the material takes place. Each of these machining processes, introduce different types of defects which can appear on different locations of the metallic surface. The defects are caused by sub-optimal parametrization of the process or by pre-existing defects of the original raw material, that were propagated to later stages. Inspection is typically performed in the final stage before product packaging and involves human judgement and expertise. Manual inspection has not been yet widely replaced by automated solutions since in many industrial applications, the performance of these systems fail to reach sufficient accuracy levels. However, even if manual inspection is the preferred method for many industries, it is still an error-prone and time consuming process.
In order to automate surface inspection in sink manufacturing, this work provides the following contributions

- We designed a sink inspection system using machine vision sensors, illuminaton and a mounting frame to secure the object for inspection.
- We created a dataset of surface defects using sink images from the production.
- We benchmark UNet variants for defect detection and examined the impact of different loss functions on the performance.

2 RELATED WORK

In recent literature, defect inspection in metallic surfaces has been studied by several works. In [11], the authors study 120 publications for automated computer vision based defect detection

systems for application in flat steel manufacturing. They identify the main challenges derived from sub-optimal imaging conditions and processing of large amount of image data, which raises the need of efficient trade-off between accuracy and complexity. The paper classifies four main catregories of defect detection methods, the statistical, the spectral, the model-based and the machine learning. Machine learning and specifically Deep Learning has been widely studied and proven to surpass traditional approaches in image processing. Therefore, latest studies focus on DL methodologies for all the computer vision tasks such as image segmentation, object detection and image classification. In [12], the authors propose CVAE network (Convolutional Variational Autoencoder) for the automated defect detection in metallic surfaces, demonstrating performance gain of 3% in F1-score adding a data augmentation step to deal with data scarcity. In another work, [13], an object detection system is proposed for the detection of surface defects on steel strips. The authors propose an improvement of the known network YOLO (You Only Look Once) in order to be able to perform in real-time with high performance. The experimental results demonstrated mean Average Precision (mAP) of 97.5% , surpassing other approaches. In [14], the authors use RetinaNet and combine classification, detection and tracking for the defects on metal surfaces from the automotive industry. They exploit the temporal coherence in consecutive frames from a camera and demonstrated performance in mAP of 76%, outperforming other methodologies that use static images. In another recent work, [15], a visual inspection system is proposed for defect detection on metal surfaces. The author developed Deformable Convolution and Concatenate Feature Pyramid Neural Network, DCCFP-Net and showed performance gain of approximately 3% compared to MobileNet [16].

In semantic segmentation tasks, UNet has been proposed in [17], initially for biomedical image segmentation. UNet is built upon the fully convolutional network approach. However, it is adjusted so that it can work efficiently with few training images, while improving segmentation precision. Since the original version, various modifications and improvements have been studied, based on UNet encoder-decoder structure [18, 19, 20]. In [21], the authors employ a weighted attention mechanism and skip connections to propose Res-UNet in order to segment precisely small thin retina vessels. In the same domain, in [22] a spatial attention mechanism and structed dropout convolutional blocks are introduced to the UNet architecture. The proposed SA-UNet achieves state-of-the-art performance on retinal vessel segmentation task. In [23], Eff-UNet is presented, which combines EfficientNet [24] as the encoder for feature extraction and UNet decoder for reconstructing the segmentation map. The proposed approach achieved 0.6276 mean Intersection over Union (mIoU) in a dataset concerning an unstructured driving environment. UNets have been widely studied in the field of medical imaging but recent works have demonstrated their efficiency also in different domains. In [25] the authors study a UNet-based semantic segmentation network for detecting defects in textured surfaces. In their work, they investigate the effectiveness of approach using weakly annotated data, showing performance of 0.6762 in mIoU. In [26] a novel approach is proposed for defect detection, based on Depth-wise Squeeze and Excitation Block-based Efficient-UNet, DSEB-EUNet. The experiments demonstrated performance of the proposed approach of 0.7180 in mIoU which surpasses other UNet approaches, including Eff-UNet. Finally, in another recent work [27], the

authors use atrous spatial pyramid pooling, ASPP, and propose S-Unet which exploits different receptive fields to improve segmentation accuracy of targets with different sizes. The proposed network outperforms UNet in precision, recall and F-score, in the industrial cases of surface defect segmentation on commutators and steel surfaces.

3 PROPOSED APPROACH

Following the recent research on defect detection using semantic segmentation approaches, in our work we compare several UNet variants. We investigate the potential of the networks to identify and localize defective regions on metallic surfaces, applied in sink-manufacturing.

3.1 UNet architecture

The UNet family of architectures follow a U-shaped network structure, which is a symmetric fully convolutional neural network. In Figure 1, the general architecture of UNets is illustrated. The architecture is based on an encoder-decoder design, where in the encoder part, the input follows a contraction path and in the decoder part, the extracted features follow an expansion path. The contraction path consists of a stack of convolutional and pooling layers, $E[1-5]$, to capture the image context while the expansion path cosists of transposed convolutions, $D[1-4]$, in order to enable the precise localization of the target regions. Between the symmetrical contraction and expansion paths the skip connections represent the operation of copying and concatenating the image features. These connections help propagate the context information from the low level features to the high level features. Since the encoder part of the UNet architecture follows the typical deep convolutional structure, modifications in the structure can be made using other networks as the encoder, including ResNet [28] and EfficientNet[24]. The ResNets employ residual blocks which solve the problem of vanishing gradient when training deeper neural networks. From ResNet family of networks, several variants have been proposed from lower to higher network complexity, namely ResNet-18, ResNet-34, ResNet-50, ResNet-101 and ResNet-152. In our work we examine ResNet-50 variant which is commonly used as backbone network, since it balances the trade-off between accuracy and speed. EfficientNets employ inverted residual blocks to reduce the network's parameters, while a scaling method is used to uniformly scale the network width, depth and resolution. In [24], eight EfficientNet models were initially examined with different complexity, starting from the less to more complex, EfficientNet-B0 to EfficientNet-B7. In our work, we examine three of these models, the versions B0, B3 and B7 in order to cover three different network complexities.

3.2 Metrics

For semantic segmentation tasks, the goal is to classify each pixel of the image based on the examined categories. In our case we have two categories, the defective and non-defective. We will use precision, recall and F1-score [29] to investigate the performance of each model. F1-score is equivalent to Dice Coefficient (DC) which is commonly used in segmentation tasks

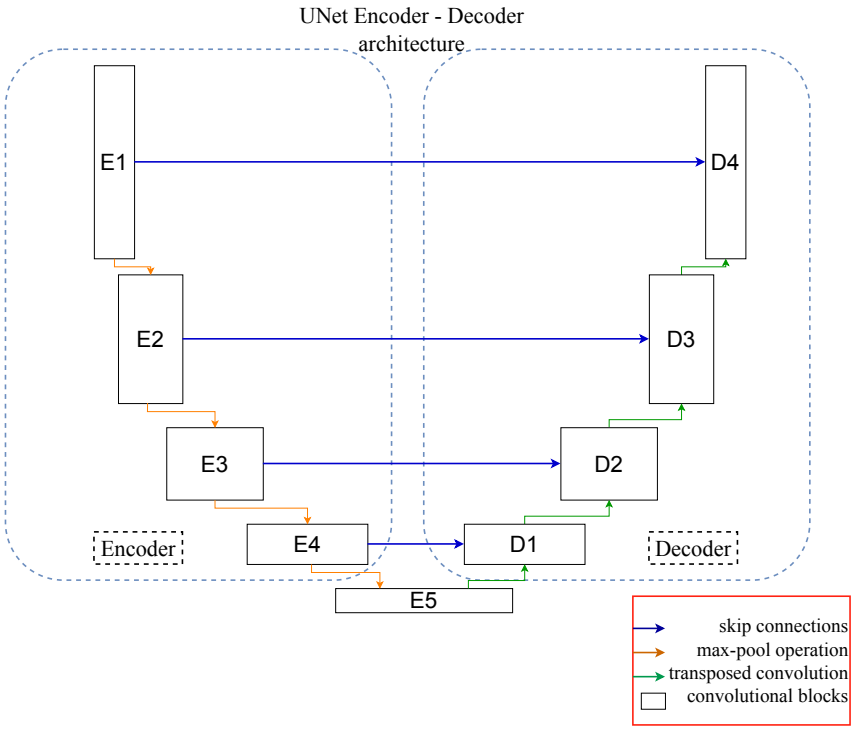


Figure 1: UNet general architecture.

and represents the similarity between two samples [30]. The higher the DC score, the larger the overlap between the prediction and the ground truth and thus the more accurate the model in localizing the defect. Additionally, we will report IoU metric which is also widely used and is given by Equation 1, where TP, FP, FN are the true positive, false positive and false negative pixels.

$$IoU = \frac{TP}{TP + FP + FN} \tag{1}$$

3.3 Data description

The machine vision system that we designed for sink inspection is illustrated in Figure 2 on the left. The system consists of two machine vision sensors, Hikrobot MV-CS200-10GC of 20 Megapixels. For the illumination of the environment, we used two LED luminaires in order to acquire the images under adequate light. The sensors were mounted inside a closed inspection booth, so that the reflections from external lights are limited. In order to create the image dataset, the sink samples were placed in the inspection booth and high resolution images were captured from the samples. The size of the original images were 5472×3648. Due to limited amount of defective samples, we followed an augmentation procedure to create sufficient amount of data in order to efficiently train the DL networks. We initially annotated

the defects of the dataset using open-source software. Then, from the defective regions, we extracted image patches of size 384×384, cropped from the original images of sink samples that were captured with high resolution. The cropping of the patches was performed using a shifting window scheme on the original image with overlapping of 25% between consecutive patches. Indicative examples of the original images are shown in Figure 2 (right) while in Figure 3 we see indicative examples of the extracted image patches, with surface defects of different types and shapes, including scratches, dents and bumps. Each image patch includes defective (foreground) and non-defective (background) pixels and we aim at examining how UNet variants can identify and localize the defective regions.

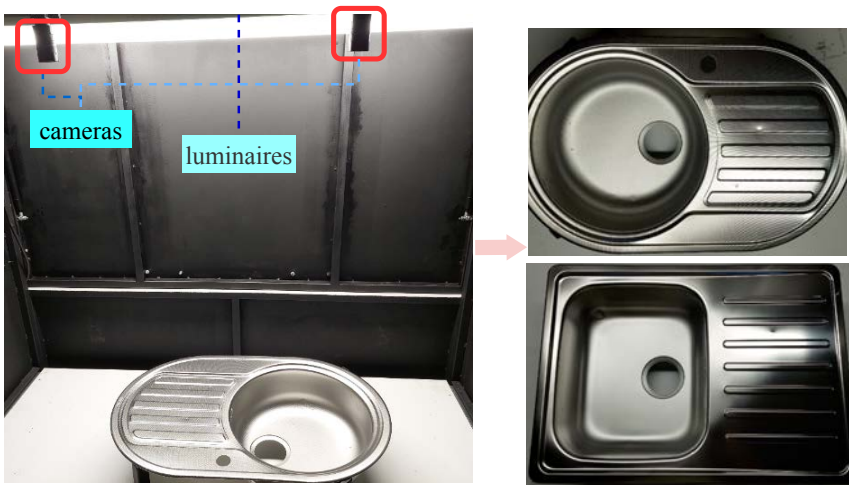


Figure 2: The inspection system comprises machine vision sensors and luminaires (left). On right, sink samples as captured by the system.

The final dataset comprises of 1494 image patches for training, 117 for validation and 670 for testing. The image patches for each of the train, validation and test dataset correspond to different sink samples. Thus, the test dataset is used during evalution in order to investigate the generalization capability of the trained network. In order to avoid overfitting problems when training the network and increase the diversity of the training data, we applied common image augmentation techniques involving color transformations and geometric transformations.

3.4 Network training

For the training of the network, we trained for 50 epochs and for each epoch the performance of the trained model on the validation dataset was monitored. After each period of 7 consecutive epochs, if the validation performance of the model was not improved, an early stopping was applied. After the training, the performance of the final model on the test dataset was recorded for the evalution. The network was trained using Adam optimizer with learning rate of 0.0001. The experiments were performed using PyTorch library and the NVIDIA RTX 3080Ti graphics card. As mentioned in Section 3.3, the original images were annotated and then cropped using

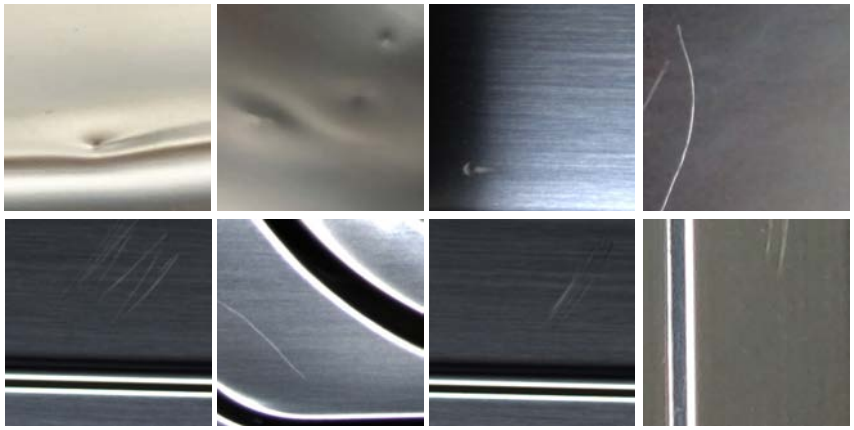


Figure 3: The dataset includes image patches with different types of surface defects.

a shifting window scheme to create sufficient amount of training data. During the training of the network, the color images are fed to the input of the network while the segmentation mask is used as the target. In Figure 4, we can see an image patch that is fed to the network and the corresponding segmentation mask that is used for the calculation of the loss function. For the loss function we investigated the performance impact of three different loss functions, the binary cross entropy loss L_{bce}, Equation 2, the dice loss L_{dice} , Equation 3 and the weighted combination of the two L_{wbd}, Equation 4. In the equations, $y_i \in \{0, 1\}$ denotes the ground truth label while $p_i \in [0, 1]$ is the predicted probability and n is the number of image pixels. In Equation 4, L_{wbd} is a weighted combination of L_{bce} and L_{dice} and w_1, w_2 represent weighting factors for each term. In our experiments, we concluded by experimentation that the values $w_1 = 0.05$ and $w_2 = 0.95$ provided performance gain for most of the experiments.

$$L_{bce} = -\sum_{i=1}^{n} (y_i log(p_i) + (1 - y_i) log(1 - p_i))) \tag{2}$$

$$L_{dice} = 1 - \frac{2\sum_{i=1}^{n} p_i y_i}{\sum_{i=1}^{n} p_i^2 + \sum_{i=1}^{n} y_i^2} \tag{3}$$

$$L_{wbd} = w_1 L_{bce} + w_2 L_{dice} \tag{4}$$

4 EXPERIMENTAL RESULTS

As mentioned in Section 3.2, we use Precision(PR), Recall (REC), F1-score (F1) and IoU to measure the performance of the compared models. For each model, the performance was calculated using the actual segmentation mask which was compared to the predicted mask, in pixel level. The goal of the evaluation is to examine the capability of the model to localize

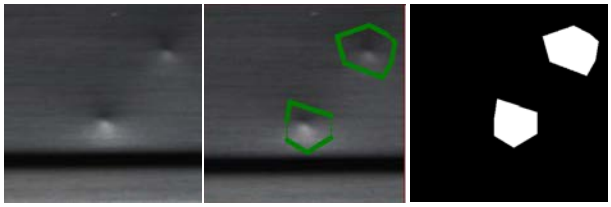


Figure 4: The defective regions are annotated (left and middle) and the binary segmentation mask (right) is used as the network's target.

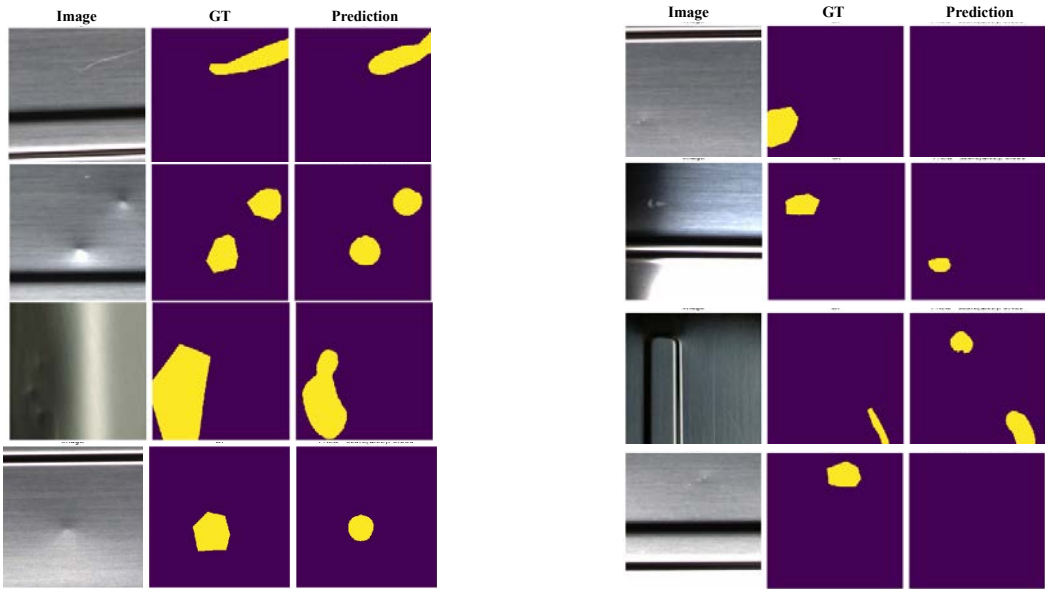
accurately the defective regions and distinguish them from the background, which corresponds to non defective regions. We performed a total of 15 experimental runs for all the combinations of the models with the examined loss functions. For each of the experiment, we report in Table 1 the performance of the model on the test set. In all the experimental runs, we use the same network configuration parameters and same augmentation steps during training.

Table 1: Segmentation results on test dataset

Model	Loss	PR	REC	F1	IoU
UNet	L_{bce}	0.469	0.066	0.116	0.061
	L_{dice}	0.301	0.142	0.193	0.107
	L_{wbd}	0.184	0.268	0.218	0.122
Res50-UNet	L_{bce}	0.886	0.421	0.571	0.399
	L_{dice}	0.604	0.518	0.558	0.387
	L_{wbd}	0.647	0.587	0.615	0.444
EffUNet-B0	L_{bce}	0.779	0.546	0.642	0.473
	L_{dice}	0.644	0.685	0.664	0.497
	L_{wbd}	0.672	0.653	0.662	0.495
EffUNet-B3	L_{bce}	0.843	0.495	0.624	0.453
	L_{dice}	0.802	0.574	0.669	0.503
	L_{wbd}	0.770	0.618	0.686	0.522
EffUNet-B7	L_{bce}	0.831	0.627	0.715	0.556
	L_{dice}	0.734	0.690	0.712	0.553
	L_{wbd}	0.721	0.717	**0.719**	**0.561**

From the results, firstly we notice that for the original UNet, the performance is poor and it indicates that the model failed to converge. This can be justified by the fact that the network did not use pre-trained weights and was trained from scratch. Adding ResNet and EfficientNet, pre-trained on ImageNet [31], as backbones in UNet architecture, helps as expected with the convergence of the training and significantly boost performance. Firstly, we notice that even using the light version of EfficientNets, B0, as backbone, we get approximately 5% improvement in both IoU and F1-score, using L_{wbd} loss function, compared to using ResNet-50. We also see that using the dice loss function provides better results than binary cross entropy when

training EffUNet-B0 and B3. However, using the weighted combination of both, we achieved improvement when training UNet, Res50UNet, EffUNet-B3 and B7. For example, for EffUnet-B3, using the weighted loss function improves f1 score by 1.7% and IoU by 1.9% while for EffUNet-B7, it improves F1-score by 0.4% and IoU by 0.5%. Finally, we see in bold, that EffUnet-B7 model achieves the best performance, improving F1-score by 3.3% compared to the next best examined model EffUNet-B3. In Figure 5 we see the qualitative results from the test dataset, where on the left column of each 4x3 image grid, there is the input image, the middle column is the ground truth segmentation mask, denoted as 'GT' and the right column is the mask predicted by the model. On the left image Figure 5a, we see cases where the model has predicted accurately the defective region with dice coefficient greater than 0.7, which corresponds to high overlap between the actual defect and the prediction. On the right, in Figure 5b, we see cases where the model was not able to predict the defects.



(a) (b)

Figure 5: Qualitative results showing indicative accurate (a) and inaccurate(b) predictions.

5 CONCLUSIONS AND FUTURE WORK

In this work, we examined a system that has the potential to automate surface inspection in metallic surfaces. The proposed system uses machine vision sensors and an illumination system and is validated in a use case of inspection in sink manufacturing. Firstly, we used the system to capture multiple images from sink samples and created a dataset with various surface defects. We examined UNet variants in their ability to localize precisely the defective regions. Finally, we investigated the impact of different loss functions on the models' performance. We

concluded that EfficientNet-B7 provided the best performance in terms of F1-score and IoU, using a weighted combination of binary cross entropy and dice loss.
The final results indicate that further improvements should be examined in terms of both DL methods and image quality. In future work, we will examine DL methods to automatically control the effect of illumination in the metallic surfaces and alleviate image quality problems derived from specularity. Another future step, is to enhance the dataset with more images from the production and examine the capability of DL methods to classify the defective regions based on the separate defect categories.

ACKNOWLEDGEMENT

This work has been co-funded by the European Union and the General Secretariat for Research and Innovation, Ministry of Development & Investments, under Project PYRAMIS 4.0/T2EDK-04427. This work was also supported by the European Commission through Project OPTIMAI funded by the European Union H2020 programme under Grant 958264. The opinions expressed in this article are those of the authors and do not necessarily reflect the views of the European Commission.

REFERENCES

[1] E. I. Papageorgiou, T. Theodosiou, G. Margetis, N. Dimitriou, P. Charalampous, D. Tzovaras, and I. Samakovlis, "Short survey of artificial intelligent technologies for defect detection in manufacturing," in *2021 12th International Conference on Information, Intelligence, Systems and Applications (IISA)*. IEEE, Jul. 2021.

[2] Z. Zhang, G. Wen, and S. Chen, "Weld image deep learning-based on-line defects detection using convolutional neural networks for al alloy in robotic arc welding," *Journal of Manufacturing Processes*, vol. 45, pp. 208–216, Sep. 2019.

[3] F. Nikolic, I. Stajduhar, and M. Canadija, "Casting defects detection in aluminum alloys using deep learning: a classification approach," *International Journal of Metalcasting*, vol. 17, no. 1, pp. 386–398, Mar. 2022.

[4] W. Cao, Q. Liu, and Z. He, "Review of pavement defect detection methods," *IEEE Access*, vol. 8, pp. 14 531–14 544, 2020.

[5] J.-F. Jing, H. Ma, and H.-H. Zhang, "Automatic fabric defect detection using a deep convolutional neural network," *Coloration Technology*, vol. 135, no. 3, pp. 213–223, Mar. 2019.

[6] O. Badmos, A. Kopp, T. Bernthaler, and G. Schneider, "Image-based defect detection in lithium-ion battery electrode using convolutional neural networks," *Journal of Intelligent Manufacturing*, vol. 31, no. 4, pp. 885–897, Aug. 2019.

[7] T. Kotsiopoulos, L. Leontaris, N. Dimitriou, D. Ioannidis, F. Oliveira, J. Sacramento, S. Amanatiadis, G. Karagiannis, K. Votis, D. Tzovaras, and P. Sarigiannidis, "Deep multisensorial data analysis for production monitoring in hard metal industry," *The International Journal of Advanced Manufacturing Technology*, vol. 115, no. 3, pp. 823–836, Oct. 2020.

[8] S. Bounareli, I. Kleitsiotis, L. Leontaris, N. Dimitriou, A. Pilalitou, N. Valmantonis, E. Pachos, K. Votis, and D. Tzovaras, "An integrated system for automated 3d visualization and monitoring of vehicles," *The International Journal of Advanced Manufacturing Technology*, vol. 111, no. 5-6, pp. 1797–1809, Oct. 2020.

[9] L. Leontaris, N. Dimitriou, D. Ioannidis, K. Votis, D. Tzovaras, and E. Papageorgiou, "An autonomous illumination system for vehicle documentation based on deep reinforcement learning," *IEEE Access*, vol. 9, pp. 75 336–75 348, 2021.

[10] N. Dimitriou, L. Leontaris, T. Vafeiadis, D. Ioannidis, T. Wotherspoon, G. Tinker, and D. Tzovaras, "Fault diagnosis in microelectronics attachment via deep learning analysis of 3-d laser scans," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 7, pp. 5748–5757, Jul. 2020.

[11] Q. Luo, X. Fang, L. Liu, C. Yang, and Y. Sun, "Automated visual defect detection for flat steel surface: A survey," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 3, pp. 626–644, Mar. 2020.

[12] J. P. Yun, W. C. Shin, G. Koo, M. S. Kim, C. Lee, and S. J. Lee, "Automated defect inspection system for metal surfaces based on deep learning and data augmentation," *Journal of Manufacturing Systems*, vol. 55, pp. 317–324, Apr. 2020.

[13] J. Li, Z. Su, J. Geng, and Y. Yin, "Real-time detection of steel strip surface defects based on improved YOLO detection network," *IFAC-PapersOnLine*, vol. 51, no. 21, pp. 76–81, 2018.

[14] S. B. Block, R. D. da Silva, L. B. Dorini, and R. Minetto, "Inspection of imprint defects in stamped metal surfaces using deep learning and tracking," *IEEE Transactions on Industrial Electronics*, vol. 68, no. 5, pp. 4498–4507, May 2021.

[15] Z. Liu, B. Yang, G. Duan, and J. Tan, "Visual defect inspection of metal part surface via deformable convolution and concatenate feature pyramid neural networks," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 12, pp. 9681–9694, Dec. 2020.

[16] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," 2017.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science*. Springer International Publishing, 2015, pp. 234–241.

[18] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, and J. Wu, "UNet 3+: A full-scale connected UNet for medical image segmentation," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2020.

[19] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-denseunet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Transactions on Medical Imaging*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[20] L. Cheng, J. Yi, A. Chen, and Y. Zhang, "Fabric defect detection based on separate convolutional UNet," *Multimedia Tools and Applications*, vol. 82, no. 2, pp. 3101–3122, Jul. 2022.

[21] X. Xiao, S. Lian, Z. Luo, and S. Li, "Weighted res-UNet for high-quality retina vessel segmentation," in *2018 9th International Conference on Information Technology in Medicine and Education (ITME)*. IEEE, Oct. 2018.

[22] C. Guo, M. Szemenyei, Y. Yi, W. Wang, B. Chen, and C. Fan, "SA-UNet: Spatial attention u-net for retinal vessel segmentation," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, Jan. 2021.

[23] B. Baheti, S. Innani, S. Gajre, and S. Talbar, "Eff-UNet: A novel architecture for semantic segmentation in unstructured environment," in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE, Jun. 2020.

[24] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," 2019.

[25] N. Enshaei, S. Ahmad, and F. Naderkhani, "Automated detection of textured-surface defects using UNet-based semantic segmentation network," in *2020 IEEE International Conference on Prognostics and Health Management (ICPHM)*. IEEE, Jun. 2020.

[26] H. Üzen, M. Turkoglu, M. Aslan, and D. Hanbay, "Depth-wise squeeze and excitation block-based efficient-unet model for surface defect detection," *The Visual Computer*, Mar. 2022.

[27] D. Wang and Y. Liu, "An improved neural network based on UNet for surface defect segmentation," in *3D Imaging Technologies—Multidimensional Signal Processing and Deep Learning*. Springer Singapore, 2021, pp. 27–33.

[28] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015.

[29] C. Goutte and E. Gaussier, "A probabilistic interpretation of precision, recall and f-score, with implication for evaluation," in *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2005, pp. 345–359.

[30] S. Jadon, "A survey of loss functions for semantic segmentation," in *2020 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*. IEEE, Oct. 2020.

[31] Y. Gao, J. Lin, J. Xie, and Z. Ning, "A real-time defect detection method for digital signal processing of industrial inspection applications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 5, pp. 3450–3459, May 2021.