DOI: 10.53375/ijecer.2023.320

RIAWELC: A Novel Dataset of Radiographic Images for Automatic Weld Defects Classification

Benito Totino[1], Fanny Spagnolo[1], Stefania Perri[2,*]
[1]Dept. of Informatics, Modeling, Electronics and Systems Engineering, University of Calabria, Rende, Italy
Email: benito.totino@gmail.com, f.spagnolo@dimes.unical.it
[2]Dept. of Mechanical, Energy and Management Engineering, University of Calabria, Rende, Italy
Email: s.perri@unical.it
[*] *Corresponding author*

Abstract: **In the last few years, extracting, analyzing and classifying welding defects in radiographic images received a great deal of attention in several industry manufacturing. Nowadays, computer vision affords considerable accuracy in many practical applications, but making automatic processes approachable also in this field is still a challenge. As an example, Convolutional Neural Networks (CNNs) are widely recognized as efficient and accurate classification structures, but, due to the limited availability of specific datasets, training a CNN to classify welding defects is not trivial. This paper presents a new dataset collecting 24,407 radiographic images related to several classes of welding defects: lack of penetration, cracks, porosity and no defect. The proposed dataset of welding defects in radiographic images is released freely to the research community. As an example of application, the dataset has been used to train a customized version of the SqueezeNet CNN obtaining a test accuracy higher than 93%.**

Keywords: **X-ray image dataset, weld defects, CNNs, Industrial IoT.**

I. INTRODUCTION

With the ever-growing progress of the Internet of Things (IoT) technologies for Industry 4.0, the introduction of automated inspection systems has become crucial in almost all manufacturing processes [1]. In this context, there is the need for easy-to-integrate digital solutions that can be efficiently combined to the existing productive apparatus without requiring any complex re-engineering. Moreover, such automated systems must comply with the timing and geometrical constraints imposed by the application [2].

Welding is one of the main processes used in almost all industry manufacturing and most of the products we use daily could not be realized without welds. Obviously, in any manufacturing process making precise and reliable welding is mandatory, since this directly impacts on the quality of final products. For this reason, several efforts have been spent in the past to achieve the correct welding process. Nevertheless, as it is well known, it is impossible to make a defect free welding. Therefore, inspection methods, based either on destructive or on non-destructive tests, are typically exploited to yield reliable and high-quality products. Non-destructive tests (NDT) do not damage the examined product [3] and include: 1) visual inspection; 2) liquid penetrant; 3) magnetic particle; 4) eddy current; 5) ultrasonic; 6) acoustic emission; and 7) radiography. Among the aforementioned techniques, that based on radiographic images represents an attractive solution, given that it allows achieving competitive detection accuracies with a good integration level. More important, such a kind of approach can benefit of powerful artificial intelligence (AI) algorithms, like CNNs, to classify reliably welding defects [3-7].

However, enabling CNN technologies for Industry 4.0 still presents several technical challenges, ranging from the need to ensure real-time performances also on embedded computing platforms to security issues in data transmission [8]. A further important aspect is related to the collection of appropriate image benchmarks to be used for training and validating the specific CNN. Such a task is crucial to build a robust and accurate model, but it conceals several difficulties. First, source data has to be captured on-site, with sensors located in a naturally noised environment, which can introduce disturbance and distort the actual content. Second, to avoid CNN overfitting, the dataset should include a considerable volume of images under a wide variety of conditions, which can be extremely complex and costly. Finally, once the images have been collected, they must be annotated in accordance with the defect type affecting the product. While this process is typically challenging in any field, it may result particularly difficult in industrial scenarios, requiring an adequate level of expertise in recognizing different defects and failure conditions.

As it is well known, public datasets, like ImageNet [9], MNIST [10], MS-COCO [11], SVHN [12], CIFAR-10 [13], Fashion-MNIST [14], available for the training of CNNs, collect several millions of images related to humans, animals, flowers, handwritten digits, clothes, foods, house numbers, vehicles, and so on. However, due to the different application scenario and typology of source images, such datasets are not suitable to train models for radiographic welding defect classification. To the best of our knowledge, the literature includes just a few collections of radiographic images that could be used to train a CNN to classify welding defects [15], [16]. The GDXRay [15] is a large dataset including 5 groups of images to comply with different applications (i.e. *Castings*, *Welds*, *Baggage*, *Nature* and *Settings*). Unfortunately, the *Welds* category contains only 68 images. In order to make this small set suitable to train a

CNN, further manipulations, such as image cropping and manual annotation, are required [17]. Conversely, the WDXI [16] dataset consists of 13,766 X-ray images, collecting seven different types of welding defects. Nevertheless, there is not an open-source version of such a collection, which limits the research of new solutions for welding defect classification based on CNNs.

The lack of large and public datasets, that can prevent the training of CNNs without overfitting, stimulated the establishment of the novel RIAWELC dataset, freely released at https://github.com/stefyste/RIAWELC. In its first version, presented in this paper, it provides 24,407 radiographic images, of size 224×224 pixels, related to four classes of welding defects, namely lack of penetration (LP), cracks (CR), porosity (PO) and no defect (ND). Obviously, the dataset can be extended to introduce other defects of interest. The number of images collected in the RIAWELC dataset well supports methods for automatic identification and classification of welding defects, which are desirable to achieve reliable inspection and quality control. Finally, with its amount of data resources provided, it is useful also for researchers operating in this field.

As a further contribution, this paper presents an example of application based on the SqueezeNet CNN [18], properly customized to accomplish the welding defects classification. Experimental results obtained by using images from the RIAWELC dataset for both training and testing show that a classification accuracy higher than 93% can be achieved.

II. THE RIALWEC DATASET

The original X-ray weld images, captured in a real industry-manufacturing environment and digitalized in the *jpeg* format, consist of 2000×8640 8-bit pixels each. As the first operation performed on the generic image, the weld bead has been sliced by discarding the background area. Fig. 1 reports some sample cropped images obtained in this way. Then, a specific on-purpose written software routine has been used to extract from the cropped images the regions of interest. In other words, the slices where weld defects could be located, have been windowed. This step has been performed evaluating the effects of several window sizes, ranging from 32×32 to 150×150. During this operation, we noted that a small window is appropriate to pick up small and/or close defects, like PO, but it is unsuitable for large defects, such as CR and LP. On the other hand, larger windows well fit CR and LP, but they are inadequate to pick up small defects. The window size 80×80 has been selected as a good compromise to visualize clearly enough details of different weld defects.

Taking into account the characteristics of the most popular CNN models, including SqueezeNet [18], AlexNet [19], GoogleNet [20], ResNet [21], VGG16 [22], that could be customized for the welding defects classification, extracted tiles have been scaled to the 224×224-pixel size, keeping the aspect ratio unchanged, as reported in Fig. 2.

The tiles obtained in this way have been enhanced to make edges clearer and defects more evident. In order to do this, the contrast-limited adaptive histogram equalization [23] has been applied. Finally, the collected 24,407 enhanced images have been classified into the appropriate

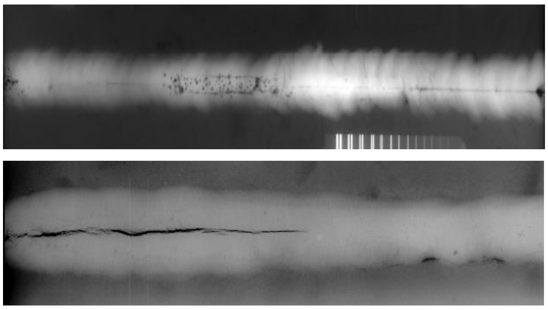


Fig. 1. Samples of cropped images.

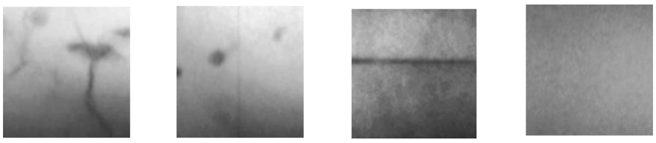


Fig. 2. Samples of 224×224 scaled tiles containing weld defects.

categories, based on the visualized defect, as summarized in Table I. The latter shows that, for each category, train, validation and test datasets have been arranged using 65%, 25% and 10% of collected images, respectively.

TABLE I. THE RIAWELC DATASET ORGANIZATION.

Defect	Total	Train	Validation	Test
CR	7635	4962	1908	765
PO	6320	4108	1580	632
LP	4452	2893	1113	446
ND	6000	3900	1500	600
Total	24407	15863	6101	2443

III. CASE STUDY: WELDING DEFECT CLASSIFICATION THROUGH RIALWEC AND A CUSTOM CNN

A. CNN model

In order to test the suitability of our dataset, several popular CNNs have been preliminarily examined. Table II summarizes the main characteristics in terms of: 1) the Top-5 error (i.e. the percentage of failures for which the proper class is not included in the top five guesses) evaluated on the ImageNet [9] dataset; 2) the number of cascaded Convolutional (Conv) or Fire layers, and their filters sizes; 3) the number of cascaded Fully Connected (FC) or Average Global Pooling (AGP) layers; 4) the total amount of trainable parameters and performed Multiply-Accumulate (MAC) operations. Table II clearly shows that the lower the Top-5 error, the more complex the CNN model. In order to limit the overall computational complexity, among the examined CNNs, we selected the SqueezeNet, which has been customized as schematized in Fig. 3.

TABLE II. CHARACTERISTICS OF STATE-OF-THE-ART CNNS.

Metrics	VGG16	GoogleNet	ResNet50	SqueezeNet V1.1
Top-5 error	7.4%	6.7%	5.3%	16.4%
Input size	224×224	224×224	224×224	224×224
# Conv/Fire Layers	13	57	53	10
Filter Sizes	3	1, 3, 5, 7	1, 3, 7	1,3
# FC Layers	3	1	1	0
# AGP Layer	0	0	0	1
Total Parameters	138M	7M	25.5M	1.24M
Total MACs	15.5G	1.43G	3.9G	349M

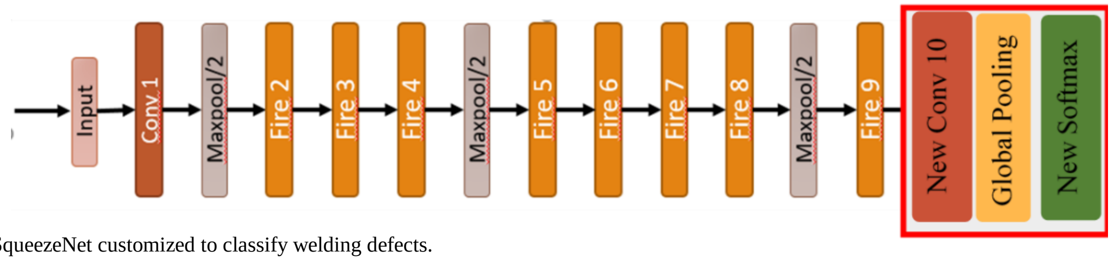
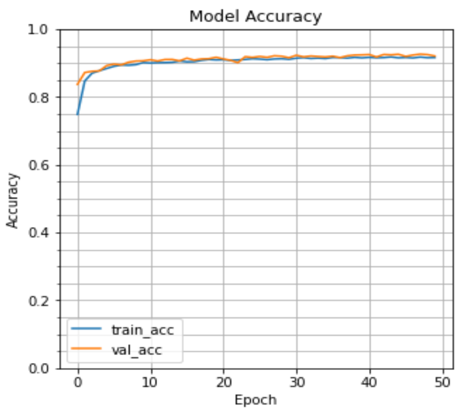


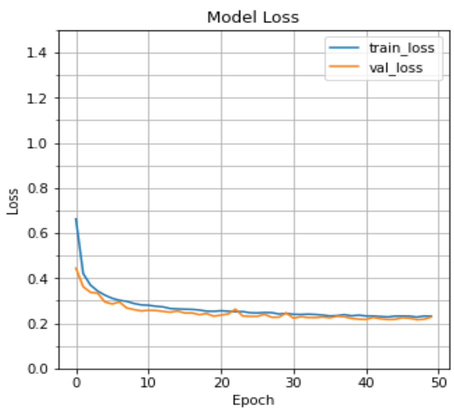
Fig. 3. The SqueezeNet customized to classify welding defects.

B. *Training and evaluation*

In order to make SqueezeNet able to classify the four selected weld defects (i.e. CR, LP, PO and ND), the transfer learning approach has been exploited. It allows applying the knowledge learned by the CNN on a large dataset, with categories unrelated to those of interest, to our own classification task. The selected baseline model has been trained on the ImageNet dataset and then it has been stripped of its last three layers that have been replaced by their customized versions on-purpose described using the Keras library [24]. Due to the introduced customization, the total number of parameters is reduced to 724548. However, thanks to the transfer learning approach, only the 2052 parameters related to the newly modified custom layers need to be re-trained. In order to do this, the early layers of the CNN model illustrated in Fig. 3 have been frozen and only the last three layers have been trained over the RIAWELC dataset. The latter has been previously normalized with the mean value, which is the same normalization adopted to train the baseline model over the ImageNet dataset. The used training settings include batch size (BS) 32, learning rate (LR) 0.001, epochs (Ep) 50 and dropout (DrO) 0.5. Moreover, we used the Adam optimization algorithm, and the CrossEntropy cost function. Results reported in Table III and Fig. 4 show that the customized SqueezeNet CNN achieves acceptable accuracy results over the RIAWELC dataset here presented. Fig. 4 also shows that the examined model approaches the maximum accuracy already on the tenth epoch. The confusion matrix and its normalization, reported in Fig. 5 and 6, confirm the good behavior of the customized CNN when examined over the test dataset.



a)



b)

Fig. 4. Accuracy and Loss behavior during the training.

TABLE III. ACCURACY RESULTS.

Metric	Accuracy
Train accuracy	92.82%
Validation accuracy	92.08%
Train Loss	0.1977
Validation loss	0.2282
Test accuracy	93.33%
Test loss	0.1904

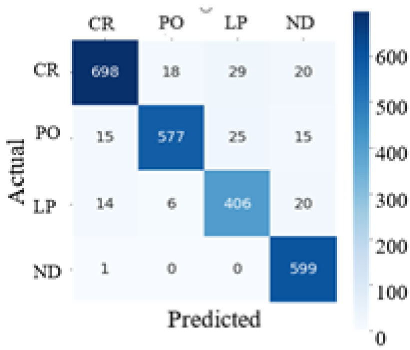


Fig. 5. Confusion matrix.

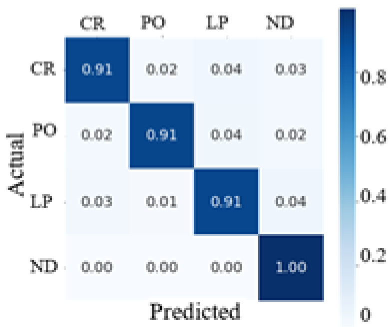


Fig. 6. Normalized confusion matrix.

Finally, Table IV collects test accuracy obtained by prior works [6], [7] and [16] using different datasets and CNN models for classification of welding defects. Due to the quite different adopted models, each leading to its own accuracy and complexity, a direct comparison with the proposed

TABLE IV. COMPARISON WITH PRIOR WORKS ON DIFFERENT DATASETS.

		[6]	[7]	[16]	This work
Dataset		GDXray	GDXray	WDXI	RIALWEC
CNN Model	# Convolution Layers	0	19	4	20
	# Fully Connected Layers	5	2	1	0
	Image size	32×32	32×32	400×400	224×224
	Filters sizes	-	3×3 1×1	5×5 3×3	3×3
#Parameters		0.7M	3.125M	0.485M	0.69M
Memory required to store the parameters [Mbytes]		2.8	12.5	1.94	2.76
#classified defects		2	5	8	4
Test Accuracy		91.84%	96.88%	46.6%	93.33%

customized model is not fair. Nevertheless, Table IV gives an interesting big picture of the state-of-the-art and allows appreciating the suitability of the proposed dataset to develop and train automatic detection CNN models. As an example, it can be observed that, when used to train and test a custom CNN model, consisting of four convolutional layers, one fully connected layer, and 1.94MBytes of parameters, the no-public WDXI dataset [16] leads to the lowest test accuracy. Conversely, training different models on images collected in the GDXray dataset [15] allows reaching promising results. However, it is worth pointing out that the models adopted in [6] and [7] are quite different than the one referred in this paper. More exactly, the AI model proposed in [6] allows just the presence/absence of defects to be distinguished. Classification of welding defects within more classes and at a higher accuracy is made possible in [7] by significantly increasing the model complexity.

It is also important to note that, since image samples from the GDXray dataset require manipulations, like cropping and defect labelling, users can exploit data augmentation to enrich the number of images as well as customize the number of classes to be identified accordingly to the application specification. Obviously, this can be achieved at the cost of an extra effort. In this perspective, the CNN model used in this work and trained through the novel RIALWEC dataset exhibits the best trade-off between accuracy and complexity. This result confirms the suitability of RIALWEC to be exploited in industrial applications where low-complex AI models are highly desirable to support welding defects classification tasks in real-time also on embedded computing platforms.

IV. CONCLUSIONS AND FUTURE WORKS

Enabling AI technologies for Industry 4.0 must still face several challenges. Among them, training robust CNN models suitable to classify welding defects with high accuracy is not trivial, because of the lack of public and large datasets containing annotated images. In order to fill this gap, this paper presented a novel dataset, named RIAWELC, consisting of 24,407 radiographic images related to several classes of welding defects. The dataset is released freely at https://github.com/stefyste/RIAWELC. Results obtained with the SqueezeNet V1.1 CNN model, demonstrated that RIAWELC is suitable for training, validate and testing CNN models customized to classify welding defects. The availability of a public ready-to-use dataset should encourage research in this direction. Further efforts will be spent in the near future to enrich the dataset with more images and defects classes, as well as to design custom CNN model and hardware/software acceleration components suitable to be integrated within embedded computing platforms like Raspberry PI or Field Programmable Gate Array (FPGA) based Sytems-on-Chips (SoCs).

ACKNOWLEDGMENT

This work is part of the project “Strumenti e Metodi Intelligenti per la DigitaL Enterprise (SMILE)”, supported by the “Ministero dello Sviluppo Economico” - PON "Imprese e Competitività" 2014-2020 FESR - Grant Number F/190084/01/X44.

REFERENCES

[1] R.S. Peres, X. Jia, J. Lee, K. Sun, A.W. Colombo, J. Barata “Industrial Artificial Intelligence in Industry 4.0 -Systematic Review, Challenges and Outlook,” *IEEE Access*, vol. 8, pp. 220121- 220139, 2020.

[2] F. Frustaci, F. Spagnolo, S. Perri, G. Cocorullo, P. Corsonello, “Robust and High-Performance Machine Vision System for Automatic Quality Inspection in Assembly Processes,” *Sensors*, vol. 22, no. 8, 2022.

[3] S. K. Dwivedi, M. Vishwakarma, Akhilesh Soni, “Advances and Researches on Non Destructive Testing: A Review,” *Proc. Int. Conf. on Materials Processing and Characterization (ICMPC)*, Hyderabad, Telangana, March 2017.

[4] S.J. Oh, M.J. Jung, C. Lim, S.C. Shin, “Automatic Detection of Welding Defects Using Faster R-CNN,” *Applied Science*, vol. 10, pp. 1-10, 2020.

[5] L. Yang, H. Jiang, “Weld defect classification in radiographic images using unified deep neural network with multi-level features,” *Journal of Intelligent Manufacturing*, vol. 32, pp. 459–469, 2021.

[6] W. Hou, Y. Wei, J. Guo, Y. Jin, C. Zhu, “Automatic Detection of Welding Defects using Deep Neural Network,” *Proc. 10th International Conference on Computer and Electrical Engineering*, IOP Conf. Series: Journal of Physics: Conf. Series 933 (2018).

[7] H. Pan, Z. Pang, Y. Wang, Y. Wang, L. Chen, “A New Image Recognition and Classification Method Combining Transfer Learning Algorithm and MobileNet Model for Welding Defects,” *IEEE Access*, vol. 8, pp. 119951-119960, 2020.

[8] A. Bécue, I. Praça, J. Gama, “Artificial intelligence, cyber-threats and Industry 4.0: challenges and opportunities,” *Artifcial Intelligence Review* (2021), vol. 54, pp. 3849–3886.

[9] J. Deng, W. Dong, R. Socher, L-J Li, K. Li, L. Fei-Fei, “Imagenet: A large-scale hierarchical image database,” *Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR)*, Miami (FL), USA, pp. 248–55, 2009.

[10] L. Deng, “The MNIST database of handwritten digit images for machine learning research,” *IEEE Signal Processing Magazine*, vol. 29, pp. 141–142, 2012.

[11] T.Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, “Microsoft COCO: Common Objects in Context,” *Proc. European Conference on Computer Vision (ECCV)*, pp. 740-755, 2014.

[12] Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A. Y. Ng, “Reading Digits in Natural Images with Unsupervised Feature Learning,” *NIPS Workshop on Deep Learning and Unsupervised Feature Learning*, Granada, Spain, pp. 1-9, 2011.

[13] A. Krizhevsky, “Learning Multiple Layers of Features from Tiny Images,” Technical Report, 2009.

[14] H. Xiao, K. Rasul, R. Vollgraf, “Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms,” *arXiv:1708.07747*, Sept. 2017.

[15] D. Mery, V. Riffoa, U. Zscherpel, G. Mondragon, I Lillo, I. Zuccar, H. Lobel, M. Carrasco, “GDXray: The database of X-ray images for nondestructive testing,” *J. Nondestructive Evaluation*, vol. 34, no. 42, 2015.

[16] W. Guo, H. Qu, L. Liang, "WDXI: The Dataset of X-ray Image for Weld Defects," *Proc. Int. Conf. on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD),* Huangshan, China, pp. 1051-1055, 2018.

[17] J.S. Kumaresan, K.S.J. Aultrin, S.S. Kumar, M.D. Anand, "Transfer Learning With CNN for Classification of Weld Defect," *IEEE Access,* vol. 9, pp. 95097-95108, 2021.

[18] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, "Squeezenet: Alexnet-Level Accuracy With 50x Fewer Parameters And <0.5MB Model Size," *Proc. Int. Conf. on Learning Representations (ICLR)*, Toulon, France, pp. 1-13, 2017.

[19] A. Krizhevsky, I. Sutskever, G. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, vol. 25, pp. 1097-1105, 2012.

[20] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," *Proc of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR),* Boston (MA), USA, pp. 1-9, 2015.

[21] K. He, X. Zhang, S. Ren, J. Sun, "Deep Residual Learning for Image Recognition," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston (MA), USA, pp. 770-778, 2015.

[22] K. Simonyan, A, Zisserman, "Very Deep Convolutional Networks For Large-Scale Image Recognition," *Proc. of Int. Conf. on Learning Representations (ICLR)*, San Diego (CA), USA, pp. 1-14, 2015.

[23] K. Zuiderveld, "Contrast Limited Adaptive Histograph Equalization," *Academic Press* Professional Graphic Gems IV, San Diego (USA), 1994.

[24] F. Chollet, & others. Keras. GitHub. Retrieved from https://github.com/fchollet/keras (2015).