Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Weld Visual Inspection based on YOLO Deep Learning Algorithm

Bodea Marius

mbodea@stm.utcluj.ro

Technical University of Cluj https://orcid.org/0000-0002-3454-3009

Short Report

Keywords: YOLO, NDE, welding imperfections, computer vision

Posted Date: March 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9212344/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

The welding industry plays a major contribution to the global GDP in the world economy. Welding processes are key enabling technologies for the most important industries, like automotive, shipbuilding, rail, aerospace, oil and gas, energy, defense, up to the construction of essential infrastructures for modern world economy and society. The paper is focused on the welding visual inspection of the welded structures using state-of-the-arts methods, based on artificial intelligence (AI) and deep learning algorithms. The recent progress of AI in object detection models has revolutionized many industries, including the non-destructive examination (NDE) of welded structures, that represent one key aspect considered in the paper. Also, the paper can be of help for other specialists in the welding area and related fields that are interested in acquiring new skills in informational technologies, particularly in object detection using YOLO algorithms.

1. Introduction

1.1. Short insight in the NDE and Welding Industry

Global manufacturing industry has overpassed $9.6 trillion in 2022 and contributed to approximately 16% of the global world GDP [1]. The welding industry accounts for about 2% of the world's GDP and it employs around 3 million people around the world [2]. According to some market research studies, the global welding market size was about $24.73 billion in 2023 and is expected to grow up to $34.18 billion by 2030 [1, 2]. In the same context, the Non-Destructive Examination Market Size was valued at USD 15.3 Billion in 2022 and is expected to reach USD 32.42 Billion by 2032, at a CAGR of 7.8% from 2022 to 2032 [3]. All these data and the estimated projections of the welding and NDE industry growth rates provide a valuable insight into the future opportunities offered by the recent progress in object detection models based on AI algorithms. Nowadays, we observe an increasing demand of high strength welded structures in various industries and applications that are manufactured by using advanced welding technologies, automation and robotics. From one perspective, this can be explained considering the large and growing deficit in qualified work force all around the world, and from a different perspective, this can be seen as a performant solution in order to eliminates human errors, to reduces production time, to improve the weld quality and to reduce overall costs, including for the NDE operations.

1.2. Object detection

Many modern industries have recognized the importance of object detection capabilities and their advantages over a broad type of activities performed until now by human operators. Internet of Things (IoT), blockchain technology, object detection, and others AI related technologies present a disruptive effect that will change the landscape of the future workforce skills, and work environment. According to a Goldman Sachs report [4], AI could replace the equivalent of 300 million full-time jobs in the next few years at global level, and about 18% of work could be automated.

As example, in a manufacturing Chinese factory about 90% of its workforce have been replaced by automation, resulting in a 250% increase in productivity and in 80% decrease in defects, according to Forbes and a MIT & Boston University report [5]. From autonomous driving of cars and trucks, surveillance operations and face detection, products brand detection, in military operations or in robotics, the identification of object of interest plays an important role for the future work characteristics. Nevertheless, today we have powerfull visual search engines that can process large amounts of data, photos, videos in different environments, industrial, social media, metaverse or even in ones that we can't imagine today.

2. YOLO Deep Learning algorithm

Computer vision is an important field of AI that uses deep machine learning algorithms to train artificial neural networks to interpret, identify and classify different objects from digital images, videos and other visual inputs, like industrial cameras, Xray equipment's [6–8], ultrasound devices etc. Computer vision can be integrated into fully automatized processes, for example in visual inspections of welded structures. The inspection systems are able to take action in real-time, when defects or quality issues are detected in industrial processes. In other words, AI enables computers to think, and computer vision enables computers to see, observe and understand (IBM).

YOLO (You Only Look Once) is an end-to-end neural network that has achieved remarkable results in object detection, compared to other similar real-time object detection algorithms available today. YOLO is able to identify and classify different types of objects and to mark them using bounding boxes, showing the confidence prediction and the class labels all at once for multiple objects.

YOLO has had an amazing evolution since 2015, when a team lead by Joseph Redmon published the first version of YOLO architecture, which was able to process image data with 45 frames per second [10]. Since then, their paper has been cited over 45287 times, and the YOLO model has gained very fast the attention of many scientists that worked in the object detection problem. One of the strongest advantages of YOLO's is the speed of data processing, the object identification and classification being made in a single stage or network.

Thus, YOLO has become one of the fastest models, best suited for applications that require image analysis in real-time, like in video surveillance, self-driving cars, augmented reality, or in our case for the weld quality inspection in real-time.

In Table 1 is presented the timeline and the main characteristics of the YOLO versions, developed after the year 2015, when the first model has been introduced [6–10].

Table 1
YOLO versions and timeline. Main characteristics.

Version & release date	CNN	mAP	Comments
v1 (2015)	Darknet-24	63.4%	Detect maximum two objects in the same grid. 63.4% mAP at 45 FPS on the VOC 2007 dataset
v2 (2016)	Darknet-19	76.8%	Detect a wider range of object classes. Faster & more accurate 76.8% mAP at 67 FPS on the VOC 2007 dataset
v3 (2018)	Darknet-53	36.2%	Anchor boxes with different scales and aspect ratios, better matching the size and shape of the objects detected. Detect objects at multiple scales.
v4 (2020)	CSPNet – 54	43.5%	Anchor boxes "k-means clustering". Cluster algorithm to group the bounding boxes into clusters for multiple objects. 43.5% mAP at 65 FPS on the COCO dataset
v5 (2020)	EfficientDet	55.8%	Improve the model's performance on imbalanced datasets. Dynamic anchor boxes.
v6 (2022)	EfficientNet-L2	52.5%	Dense anchor boxes. Higher computational efficiency. 37.5% mAP at 1187 FPS on the COCO dataset
v7 (2022)	RepConvN	56.8%	Detect a wider range of object shapes and sizes, reducing the number of false positives. The higher resolution allows the detection of smaller objects improving the overall accuracy. 56.8% mAP at 160 FPS
v8 (2023)	YOLOv8	53.9%	Ultralytics released YOLOv8. Achieved 50.2% mAP score at 1.83 msec (or 546 FPS) on the MS COCO dataset
v9 (Feb. 2024)	PGI & GELAN		Programmable gradient information (PGI) and the Generalized Efficient Layer Aggregation Network (GELAN).

It should be noted that mAP (Mean Average Precision) is an important metric used to measure the performance of the object detection models, but there is a trade between the accuracy detection or precision mAP and the time in which the image analysis is performed. Thus, the mAP is related to the number of frames per second (FPS) processed by the AI model, respectively by the YOLO. As the FPS increases, i.e. for faster processes, the YOLO model should process larger amounts of data in a shorter time (inference time), that affects the mAP detection.

COCO is the acronym for **C**ommon Objects in **Co**ntext and is a visual dataset that contains a large number of images used for machine learning projects. MS COCO is a dataset from Microsoft, which has

2.5 million labeled instances in 328k images. For comparison purposes, the OID (Open Images Dataset) from Google contains 9 million annotated images [11]. The COCO dataset is often used in research projects to benchmark algorithms and to compare the performance of real-time object detection models. The COCO dataset contains classes for object detection and tracking models, which were pre-trained for detection of 80 commonly used objects.

3. Experiments

Weld Visual Inspection is a project developed for welding applications, respectively for detection of welding imperfections using visual testing, a NDE method regulated by ISO 17637:2003. The geometric imperfections in metallic materials are classified according to EN ISO 6520-1:2007 in six groups. The Weld Visual Inspection AI model has been trained to detect and classify the following imperfections classes: cracks, craters, porosity, solid inclusions, lack of fusion, excessive convexity, burn through and spatter. Of course, there are many other weld imperfections that could be placed in the internal volume of the material, like in the weld deposit or in the heat affected zone (HAZ).

These imperfections can be detected using special methods like ultrasonic testing (UT), magnetic particle testing (MT), Eddy-current testing (ET), radiographic testing (RT) and acoustic emission testing (AT), which represent the core methods used for NDE examinations of welded structures. In this paper we focus only on imperfections that can be detected by visual testing (VT) method. The Weld Visual Inspection model presented in this paper could be used to extend the range of welding imperfections detection (i.e. for X ray image analysis), by adding new images to the dataset and by re-training the model.

In Fig. 1 are presented the general aspects of the welding imperfection classes on which the Weld Visual Inspection has been trained.

In computer vision, in particularly in the Weld Visual Inspection model one key aspect is the analyzing of the visual scene and recognizing the object of interests within an visual context that can include materials or objects with no clear boundaries, like pipes, vessels, or parts from a bigger welded structures that are not limited in shape or size. That is at least from the image source data point of view, which can present a smaller or a bigger part from the welded construction.

There are a series of challenges in the object detection problem, i.e. some objects present inherent variation in shape and size, and often there is an overlapping of the objects in the visual content (e.g. cracks with craters or inclusions) that makes the annotation and labeling operations more difficult. For example, in Fig. 1b we have a crater crack (group imperfection 104 according to ISO 6520-1:2007) that can be disposed of longitudinal (1045), transverse (1046), or radiating/star cracking (1047).

At the same time, in the same image is a crater weld, which is a shrinkage cavity produced due to shrinkage of welding pool during solidification. The welding craters are classified further in crater pipe (2024) – representing a shrinkage cavity at the end of a weld run, which was not eliminated during

subsequent weld runs and end crater pipe (2025), that is an open crater with a hole reducing the cross-section of the weld.

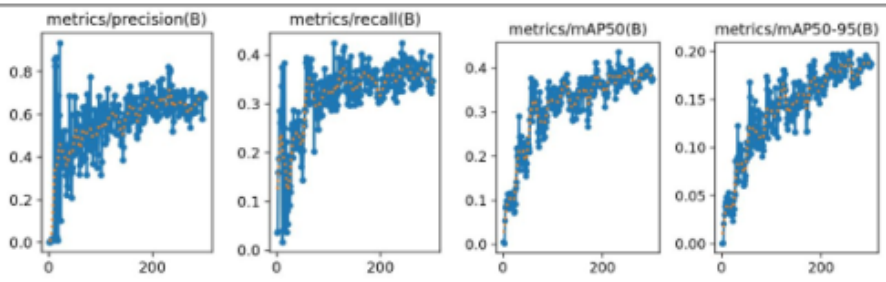
A quality dataset requires relevant high resolution of image data taken from different angles and positions, especially when the occupancy of the objects of interest in the image data could be from less than 10% up to over 80% from the source image. To these common inherent problems, we should mention the high degree of variability for lightness, contrast, saturation, exposure, and images noise.

The first version of the Weld Visual Inspection model has a dataset which contains an initial 179 images with 841 annotations and a class balance that is detailed in Fig. 3. After images augmentations operations, which include rotation (± 15°), hue (± 15°), saturation (± 25%), brightness (± 15%), exposure (± 10%) and noise up to 0.1% of pixels are resulting in a total of 455 images. These images have been divided into three sets, 399 images used for the network training, 31 images for validation and 15 images for testing. It is important to note that the images from the dataset are resized to 640 x 640 pixels, thus, to prevent the images distortion and by consequence the objects original shapes, the images should be uploaded from the very beginning in a 1:1 ratio size.

As can be seen in Fig. 4, the Weld Visual Inspection dataset has most of the images following these rules, the median width by median height of images size being 453 x 479 pixels. The gray area represents domains size that are either too tall or to wide, which can impact the train results due to higher image distortion rate.

After the training of the network using a Roboflow 3.0 model fast object detection, the performance metrics obtained are mAP 43.6%, precision 79.2% and recall 36.3% on COCO dataset. The mAP represents the Mean Average Precision and is a performance metric that is used to evaluate machine learning models. The most popular benchmarks used to assess the model object detection performance are MS COCO, ImageNET challenge, Google Open Image Challenge. The average precision (AP) achieved by class is presented in Table 2.

Table 2. Average Precision of detection by object class.

Object class	Validation	Test set	Training graphs
All objects	44 %	48 %	
Craters	57 %	26 %	
Cracks	73 %	77 %	
Lack of fusion	24 %	20 %	
Porosity	25 %	38 %	
Burn through	91 %	100 %	
Spatter	35 %	26 %	

Precision measures the quality of the model by its ability to find True Positives out of all positive predictions, which in our case are all imperfections that exist in the test images dataset. It should be

mentioned also that the performance of any object detector is evaluated through detection accuracy and inference time, which is extremely important in real time processes.
Recall (also called sensitivity) is a cuantitative metrics that is measuring the object detection model's ability to be able to find true positives out of all predictions. There is a trade-off between the two metrics, Precision and Recall. When the model has a high recall value but a low precision value, it means that the model is classifying as many negative imperfections as positive ones. On the contrary, when the model has a high precision value and a low recall value, it means that the model is classifying the welding imperfections right, but only some of them. The Weld Visual Inspection model (first version), with a mAP 43.6%, precision 79.2% and recall 36.3% has achieved quite good results using only a relatively small number of images with welding imperfections (179), see Fig. 5.

Figure 5 depicts a few examples of welding imperfection detection using the Weld Visual Inspection model built on Roboflow platform [13]. This is an opens source project that can be deployed in a series of environments like Self-Hosted Inference (a personal computer or a cloud computing platform like AWS, GCP, and Azure) or an NVIDIA Jetson or other edge devices. The NVIDIA Jetson is the world's leading platform for AI, that combines high-performance, low-power compute modules with the NVIDIA AI software stack. It’s the ideal platform for advanced robotics and other autonomous products used in various industries [14].

The metric score for Weld Visual Inspection model can be improved further by working on image data quality, by optimizing the algorithm detection, or by improving the annotation process. The quality of the dataset images is a very important aspect of the machine learning model’s performance. The images should be similar in size and ratio, also the brightness, contrast, saturation, hue and zoom level of the welding imperfections are extremely important. The tools provided by Roboflow platform allow the user to view the health status of the dataset and to search for more images that are under-represented, Fig. 3.

The optimization of the object detection algorithm is also a key factor. In Table 1 the evolution for YOLO models is provided, but that should be addressed based on the performance of the hardware used. The newest YOLO models require a lot more computation power. This led to another key performance parameter, that is the inference time, which corresponds to the object detection time expressed in milliseconds. Usualy, the accuracy of the neural network correlates with the inference time, a longer time allowing the neural network to detect more accurately the objects type and their bounding boxes.

The annotation process is an essential phase that determines the performance of the neural netwok. That requires some experience and strategy. Data annotations of the welding imperfections are typically done manually, and that can become tedious over time, especially in the case of large datasets or for small working teams. It’s obvious that the more images are in a dataset, the better it is for precision.

However, when the dataset becomes too complex or too large, there’s a lot of room for errors. Especially when some annotations can confuse the model is better to avoid the usage of those images. For instance, some cavities like slag inclusions could be vey similar in appearance to porosity or some craters.

4. Conclusions

1. The YOLO deep machine learning algorithm has revolutioned the computer vision technology in many domains. NDE technologies are among the industrial applications that will benefit from the recent progress in the field of AI applications.
2. In the light of the recent evolutions in the qualified workforce in the NDE and welding industry, is expected soon, probably before 2030, and many processes ussualy performed by human operators will be replaced by automated robots driven by AI and computer vision solutions.
3. YOLO deep learning algorithm is a very performant and accurate tool that can be used for welding imperfection detection in welded structures, not only in VT, but also in RT. These solutions can be applied in the manufacturing phase, but also in the maintenance operations for many industrial equipments and constructions.
4. Soon, large databases with different images that contain welding imperfections are expected to appear in the professional comunities, contributing to the rapid distribution of AI based applications in the industrial environment.
5. Recent evolution trends on the market capitalization show that companies like NVIDIA have surpassed big companies like Apple, Microsoft, Amazon and Google, that demonstrate the huge progress in the AI dedicated GPU hardware and software. The global AI market is expected to experience a compound annual growth rate (CAGR) of 20.2% from 2024 to 2032, potentially reaching a market value exceeding $2.74 trillion according to Nasdaq.
6. The Weld Visual Inspection application developed on the Roboflow AI platform is an open acces project and can be improved further by adding new images in the dataset and by a new session of training of the neural network.
7. The Weld Visual Inspection can be easily deployed on dedicated hardware like NVIDIA Jetson devices, allowing fast and accurate detection of welding imperfections at the moment of fabrication, cutting down the costs of inspection and repairing.

References

1. Jannik L (2024) Manufacturing Industry Statistics. Last Modified June 23, Accessed June 10, 2024. https://gitnux.org/manufacturing-industry-statistics/
2. CWB Group. Welding Industry Report 2022. Last Modified June 01 (2024) Accessed July 07, 2024. https://www.cwbgroup.org/document/id/industry-report-2022
3. Sperical Insights. Global Non-Destructive Testing Market. Analysis and Forecast 2022–2032. Last Modified April 01, 2023. Accessed July 07 (2024) https://www.sphericalinsights.com/reports/non-destructive-testing-market
4. Goldman Sachs. Generative AI could raise global GDP by 7%. Last Modified April 05, 2023. Accessed July 07 (2024) https://www.goldmansachs.com/intelligence/pages/generative-ai-could-raise-global-gdp-by-7-percent.html

5. Johnson A (2024) Forbes. Which Jobs Will AI Replace? These 4 Industries Will Be Heavily Impacted. Last Modified Mars 31, 2023. Accessed July 07. https://www.forbes.com/sites/ariannajohnson/2023/03/30/which-jobs-will-ai-replace-these-4-industries-will-be-heavily-impacted/?sh=28bd5d915957
6. Liu M, Chen Y, He L, Zhang Y, Xie J (2015) LF-YOLO: A Lighter and Faster YOLO for Weld Defect Detection of X-ray Image. *Journal of Latex Class Files, Vol. 14, No. 8*, arXiv:2110.15045v2
7. Pan K, Hu H, Gu P (2023) WD-YOLO: A More Accurate YOLO for Defect Detection in Weld X-ray Images. Sens 2023 21:8677. https://doi.org/10.3390/s23218677
8. Zuo Y, Wang J, Song J (2021) Application of YOLO Object Detection Network in Weld Surface Defect Detection. *IEEE 11th Annual Int. Conf. on CYBER Technology in Automation, Control, and Intelligent Systems, Jiaxing, China.* 10.1109/CYBER53097.2021.9588269
9. Hussain M (2023) YOLO-v1 to YOLO-v8, the Rise of YOLO and its Complementary Nature toward Digital Manufacturing and Industrial Defect Detection. Machines 11:677. https://doi.org/10.3390/machines11070677.3
10. Redmon J, Divvala S, Girshick R, Farhadi A (2015) You Only Look Once: Unified, Real-Time Object Detection. https://doi.org/10.48550/arXiv.1506.02640
11. Vidushi Meel. What is the COCO Dataset? What you need to know in 2024. Last Modified October 01 (2023) Accessed July 07, 2024. https://viso.ai/computer-vision/coco-dataset/
12. ISO 6520-1 2007. Welding and allied processes. Classification of geometric imperfections in metallic materials. Part 1: Fusion welding
13. Roboflow (2024) Everything you need to build and deploy computer vision models. Last Modified June 01, Accessed July 07, 2024. https://roboflow.com/
14. nVIDIA Developer. The NVIDIA Jetson platform. Last Modified June 01, 2024. Accessed July 07 (2024) https://developer.nvidia.com/embedded/community/support-resources

Figures

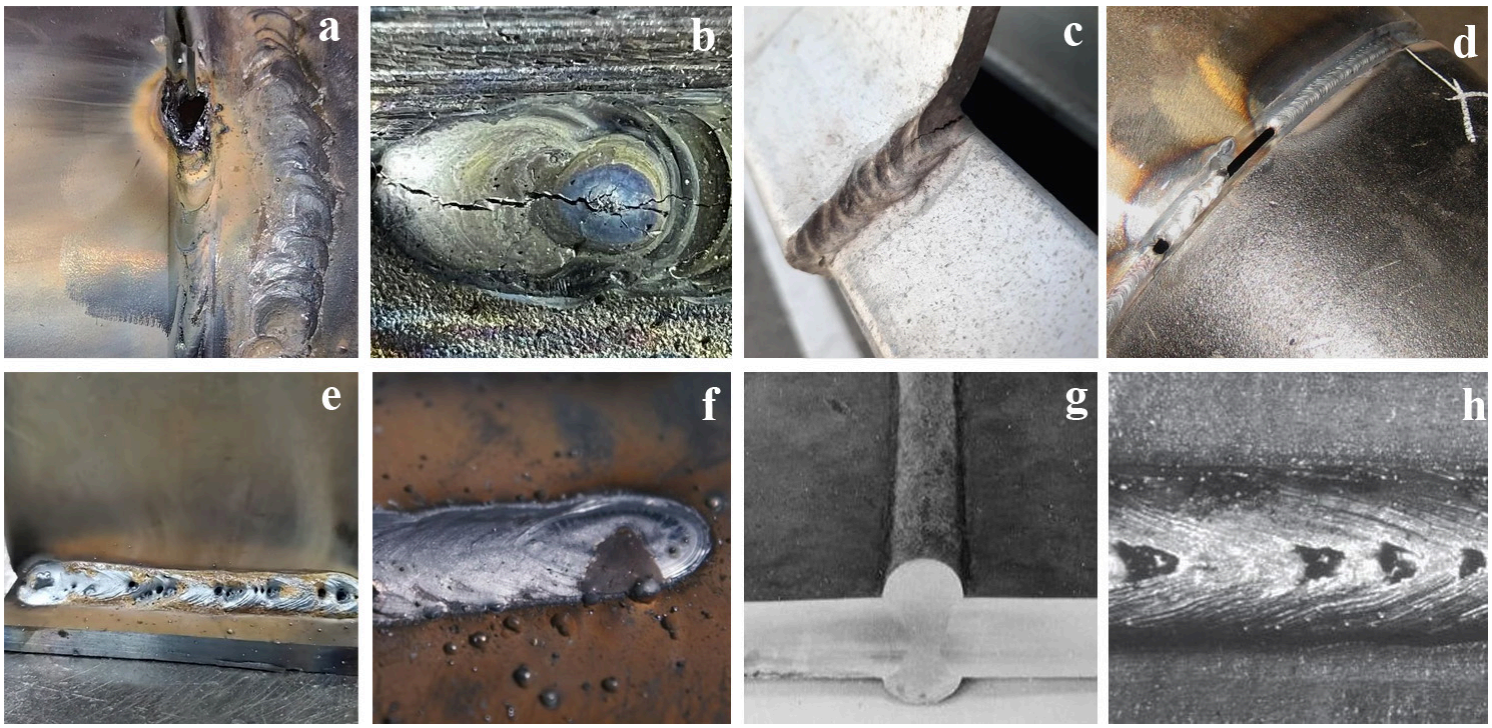


Figure 1

The classes of objects in Weld Visual Inspection model.

a) burn through, b) craters, c) cracks, d) lack of fusion, e) porosity, f) spatter, g) excessive convexity, h) solid inclusions

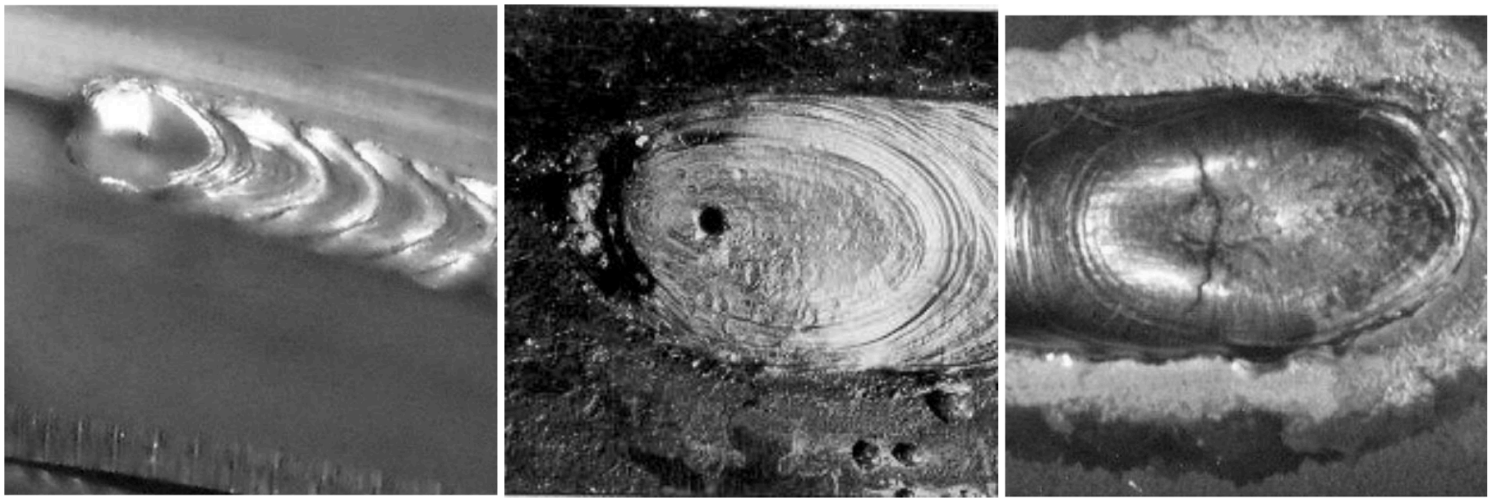


Figure 2

Welding craters, pipe (2024), open crater (2025) and transverse crack (1046) crater, ISO 6520-1:2007

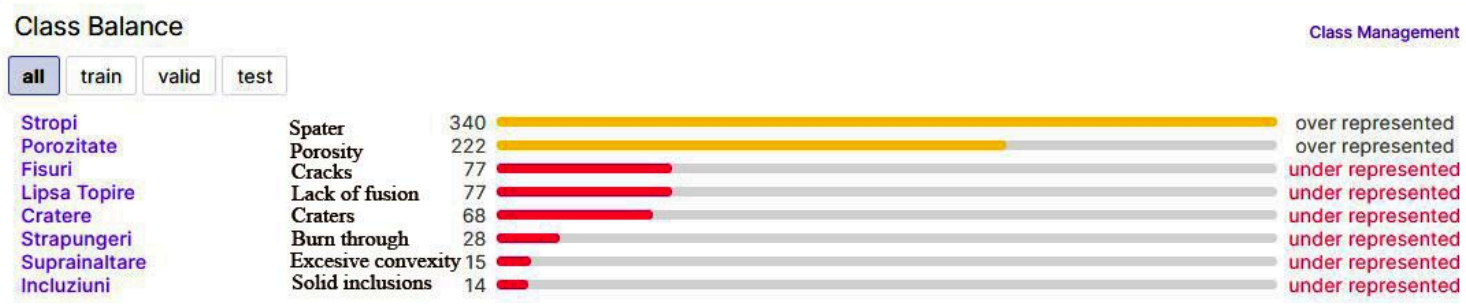


Figure 3

The Weld Visual Inspection model dataset health status.

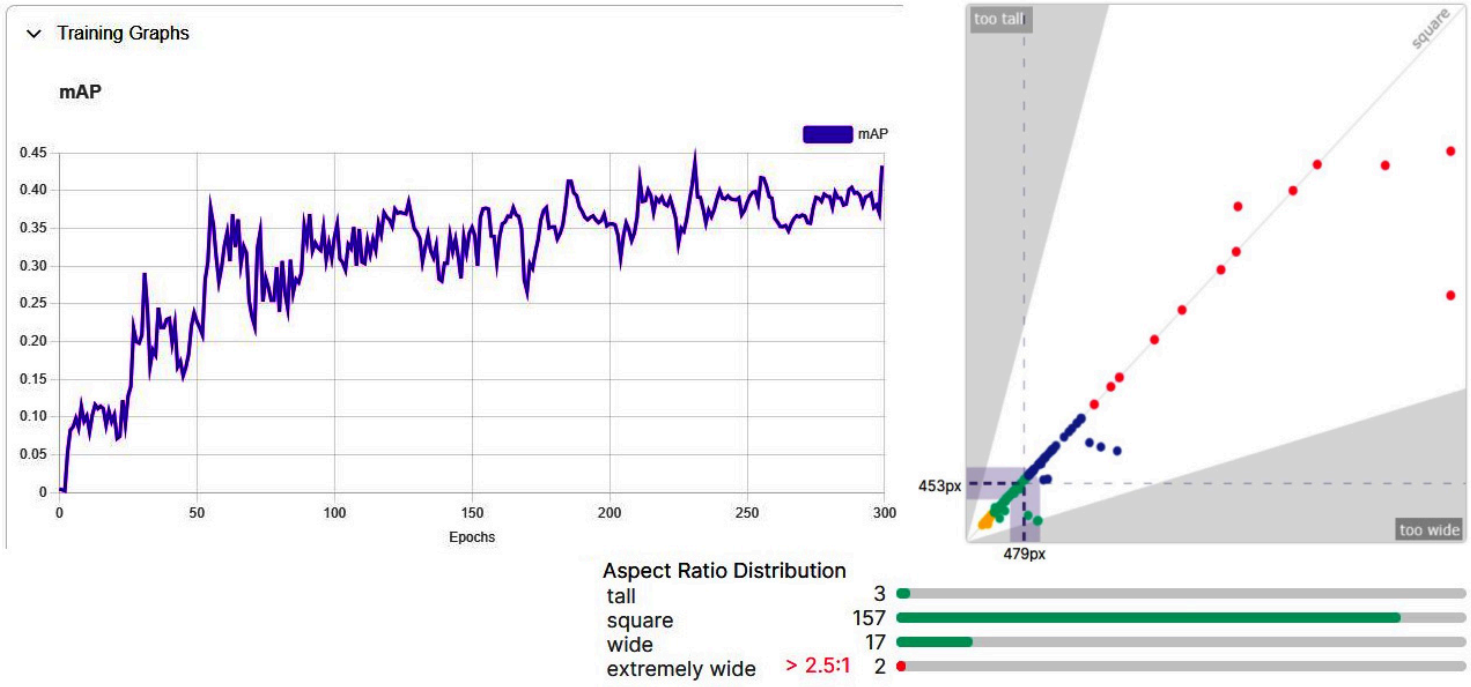


Figure 4

The Weld Visual Inspection training graph and the aspect ratio distribution of the images from dataset.

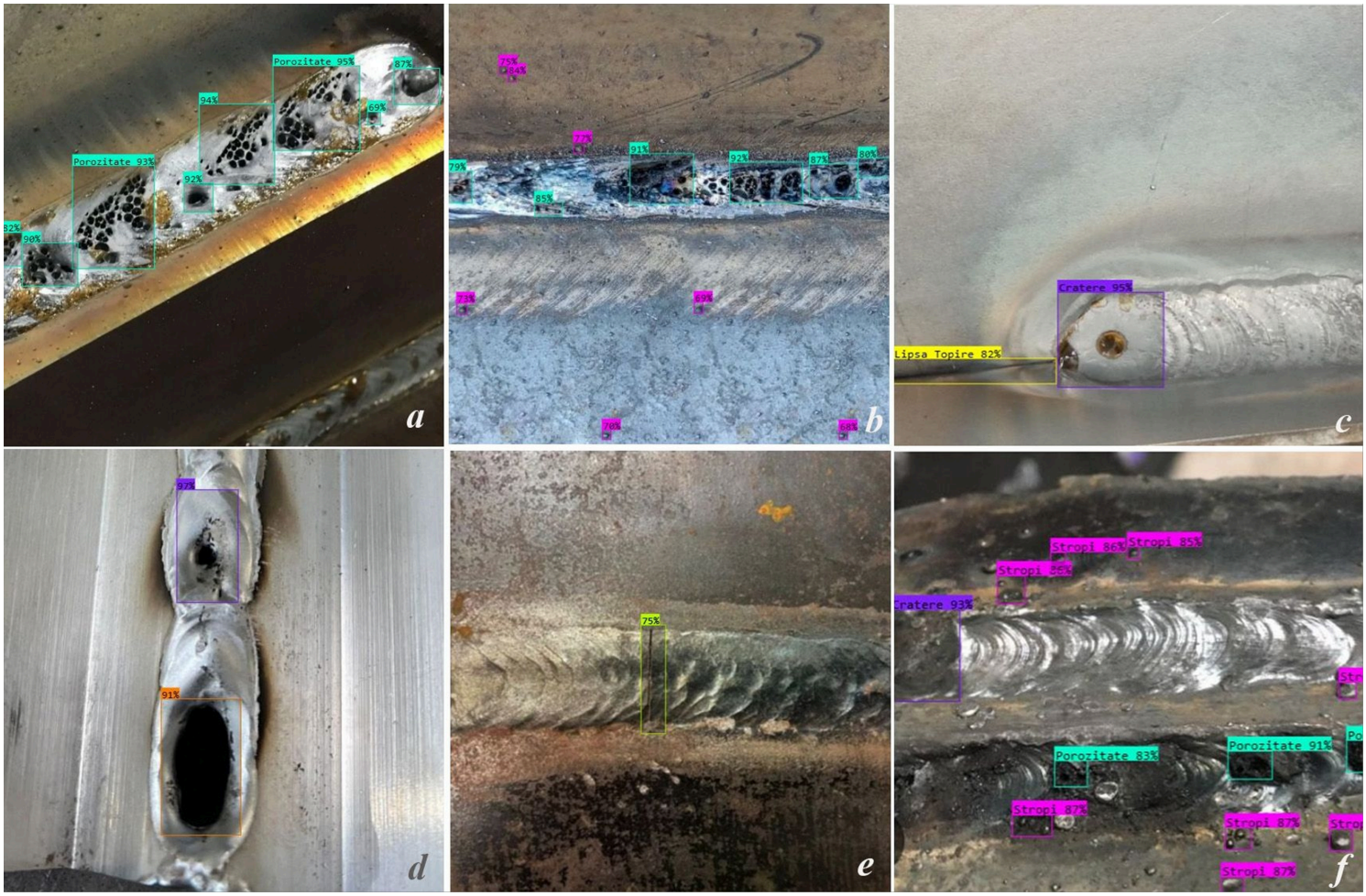


Figure 5

Welding Imperfections detection using Weld Visual Inspection model.