Dynamic Light Settings as Data Augmentations for Automated Scratch Detection

GRAVE Valentin[1], FUKUDA Osamu[2], YEOH Wen Liang[3],
OKUMURA Hiroshi[4], YAMAGUCHI Nobuhiko[5]
Graduate School of Science and Engineering, Saga University, Saga, Japan

***Abstract*—The manufacture of plastic parts requires a rigorous visual examination of its production to avoid the shipment of some that would be defective to its customers. In an attempt to ease the detection of scratches on plastic parts, the prototype of a computer-assisted visual inspection system was developed. The aim of this paper is to introduce how we explored ways to design a semi-automatic system comprising of a lamp whose orientations and intensities help in revealing irregularities on subjects that would have been missed with a unique light configuration. This process was qualified as "hardware data augmentation". The pictures collected by our system were then used to train several convolutional neural networks (YOLOv4 algorithm/architecture). Finally, the performances of their models were confronted to evaluate the effects of the different light settings, and deduce which parameters are favourable to capture datasets leading to robust defect detection systems.**

***Keywords*—Augmentation technique; deep neural network; image processing; light emission; object detection**

I. Introduction

The appearance of a product is the first and foremost characteristic that will attract, or repel, a customer. One of the key factors for an "appealing aspect" is the lack of flaws in the final product. Indeed, if defects are visible on an item, it will be considered poorly made. Therefore, it is essential for a manufacturer to carefully design its processes so that irregularities become as rare as possible.

Companies have been striving to produce quality products by implementing verification procedures in crucial stages of their production. In the plastic industry, for example, it is frequent for operators to be tasked with observing the surfaces of plastic parts under some light and comparing them with typical flawless pieces as validation standards. These pieces are moved to different positions, so the incident light may expose defects that would otherwise have remained invisible. However, visually inspecting parts is a fairly laborious and time-consuming burden for a human and the accuracy of its outcome may be unreliable. Among the few factors that can alter the judgment of an operator, their lack of experience or assiduity may certainly be the most common [1].

With the advent of artificial intelligence (AI), mostly deep learning techniques, applied to the detection of elements in pictures or video streams, industries have been trying to develop systems that would automatically notify the presence of defective items in their production lines. Nowadays, the popular method to implement such an AI is via the usage of Convolutional Neural Network (CNN) models [2], [3], [4], whose algorithms and architectures enable them to extract features from pictures (or videos) and classify their content through pattern matching.

Although current CNNs are quite efficient in detecting familiar categories of objects within pictures, they are still perfectible when used for specific tasks in specialized fields. For example, when integrated as monitoring tools on production lines of plastic or metallic parts, the scarcity of defective training data may lead to poor model performances. A substantial and diverse dataset is often required for a deep learning algorithm, such as CNNs, to produce an efficient model [5]. Moreover, another problem comes from the fact that an operator has to move and orientate the pieces in different positions to reveal flaws that would not be noticeable with a unique point of view and light incidence, since these movements are hardly ever performed automatically by an inspection system.

It is not rare for studies to explore ways that would expand limited datasets or boost the efficiency of deep learning models, but these researches generally involve software solutions revolving around image processing [6], [7]. In our case, we looked into a hardware method to enhance the quality of artificial networks' trainings through the acquisition of datasets whose elements have undergone physical data augmentations, as opposed to artificial data augmentations. These physical "data augmentations" come as fluctuations of both the brightness and the orientation of a light exposing a target (i.e. the object to be represented in the sets). This concept also aims to be useful to help models in detecting defects, as the visibility of the details of an item may vary with the intensity and the angle of incidence of its illuminating lamp. These multiple lighting conditions act as the movements an operator performs during their inspection: moving items into different positions and adjusting the brightness to facilitate their judgment, but unlike these manual inspections, our concept aims to perform these operations automatically.

II. Related Work

Researchers have investigated numerous means of performing data augmentations for both the expansion of insufficient assortments of pictures and for the creation of robust models. Our study investigates yet another novel approach to collecting and exploiting data targeting computer-assisted visual inspection of surfaces.

In deep learning, the size and diversity of a dataset are critical to creating a decent model [5], in case of a restricted set of training samples circumventing these limitations may be required. Therefore, resorting to data augmentation is a

common preprocessing step in deep learning [6], [7] (geometric transformation, noise addition, mix of images, etc.). Granted that the affinity of the new samples estimates them close to the originals, this artificial inflation of a dataset's population permits reducing the risks of a model overfit while improving its feature extraction capabilities, thus enhancing its adaptability to new data [8].

Data augmentation involving synthetic manipulations of the illumination of a picture is not unusual, whether it be by modifying their "general" brightness [9], or by adding a "synthetic" light source and shifting its position [10]. However, the majority of previous works revolve around software augmentation to generate their data, which often results in unnatural representations that may not be relevant to the models.

Moreover, some softwares and appliances aiding in the inspection of plastic parts already exist, such as the SavvyInspector™ [11] from Savvy Optics Corp., but they often do not include automatic image processing technologies nor consider different lighting configurations. Several studies also aimed at facilitating the examination of surfaces, but their experiments often did not take the orientation of the light into consideration [12], [13], [14].

Our study focused on the creation of a simple system that made possible the acquisition of data that are physically "augmented" through the different light settings used during their capture. This method is useful to generate specimens that are representative of occurrences a model will face in a live setting. Pictures of defective plastic parts were collected to train models, and the performances of these models on common test samples have subsequently been compared.

III. Image Capturing System

In current industrial environments, visual inspections necessitating rigorous examinations are still predominantly performed manually. So as to expose as many potential defects as possible on products, the common procedure is to scrutinise them under multiple lighting configurations. Operators are required to adjust the positions of the objects' surfaces, such as in Fig. 1, so the angles of incidence of the rays emitted by the light source can reveal deficiencies [15]. They may also modify the brightness of the source, since it has an obvious influence on the global visibility of the subjects they observe. Nevertheless, an equilibrium is to be found: while a dim light makes it difficult to distinguish details -if anything-, a bright light can also alter vision by flooding a defective area, covering its flaws with interfering glares.

In order to observe the effects distinct illumination settings of samples have on the efficiency of a CNN model's training, we have developed a semi-automatic device equipped with a light that automatically rotates around a subject while taking pictures as it reaches a designated set of positions. Its diagram is illustrated in Fig. 2 and its prototype is shown in Fig. 3. However, the current system version still needs the user to change the intensity of the light manually.

This system enables the capture of pictures with subjects illuminated from multiple angles and with different lighting brightness. The current prototype allows 256 levels of luminosity and has an amplitude of about 270°, since the support

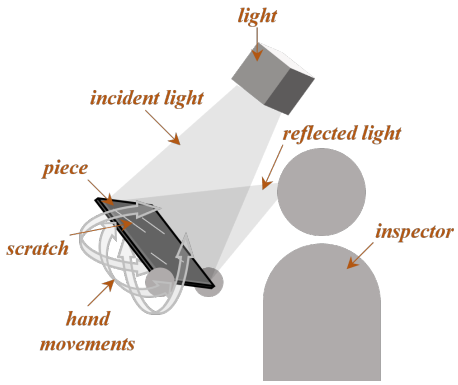

Fig. 1. Diagram of a manual inspection

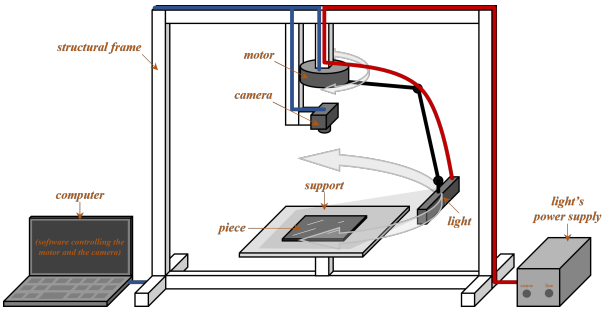

Fig. 2. Diagram of our automatic inspection setup

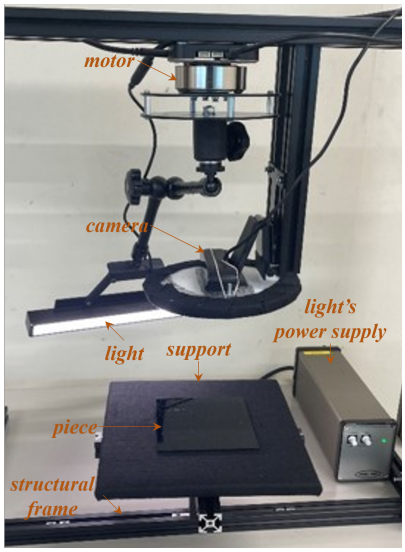

Fig. 3. View of the actual prototype

of the camera makes a full revolution impossible (as it gets in the way of the arm holding the bar light as it rotates). Because the target observed persists over a given set of light configurations, we qualified the process by which the variety of images was captured using our system as "hardware data augmentations", as opposed to a regular collection of pictures that would have been taken in a unique light setting. The augmentations thus engendered by our system took various forms: appearances/disappearances of scratches, variations of their shapes as well as shifts of their colors (or that of their supports). Several images were thereby collected to train multiple models on data having different light exposures. Their incorporations into datasets are detailed further in a subsequent section.

A. Hardware Components

Our device measures 50 by 50 by 50cm. Its constituting elements, visible on the photograph in Fig. 3, are the following:

- A camera that collects shots of plastic pieces. Created by Logitech, the C920 is a 1080p webcam that permits a stream of 30 frames per second.
- An electrical motor that orientates the light. The KM-1U is manufactured by Keigan Motor. It can be controlled programmatically and has a step resolution of about 0.05°.
- A light that illuminates the subjects that are observed. This light bar, model TLB245x25-22WD-4, is made by Aitec System. It can accept up to 24V and 0.3A, which results in a maximum power of 7.2W.
- An adjustable power supply that regulates the light intensity. The TPDP1B-2450NCW is made by Aitec System. It has 256 (16*16) gradations and it can output up to 24V±0.5V and 1.85A.

Thanks to its 256 gradations, the power supply made it possible to fine-tune the luminosity level of the light. Since the latter can accept up to 7.2W, each gradation of the power supply represents a fluctuation of 0.028125W (7.2W / 256).

Finally, to run the software that operates our prototype and to carry the CNN trainings, we used a laptop whose operating system was Windows 10 and which included an Intel i7-8750H @2.20GHz CPU, an NVIDIA GTX 1060 GPU and 16GB of RAM.

B. Operating Software

A custom software had to be developed to enable a user to control the acquisition device we created. Considering that the operating system of our working station was Windows, we thought that designing a program in C# on Visual Studio was the most practical. The User Interface (UI) of the current version is shown in Fig. 4. This UI is divided into three main sections:

- **Configuration**: in this column are listed the elements related to the establishment of connections with the external peripherals (motor and camera) and the creation of the base file structure where captures will be stored.

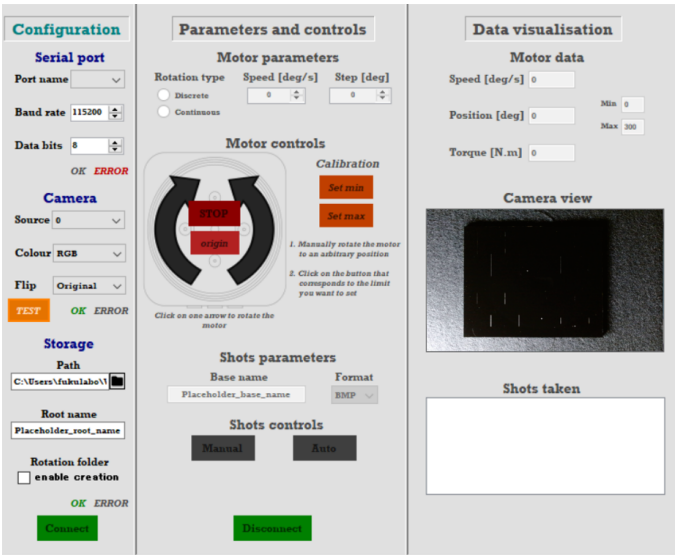

Fig. 4. Interface of our image acquisition software

- **Parameters and controls**: this column contains the components that enable the user to parameterize the motor, select how the captures are to be taken (automatically or manually) and control the motor manually (if desired).
- **Data visualisation**: data coming from the system (motor and camera) are shown within this space, so the user can monitor the acquisition process.

This program allowed us to capture a lot of pictures with ease and speed. Indeed, within the **Parameters and controls** division, it is possible to specify the angular spacing appropriate between two successive captures and to define the movement of the camera as "automatic". However, as stated previously, the current prototype necessitates the intensity of the light to be set manually. Despite its resulting semi-automatic nature, the device still fulfils the incommodious task of rotating the light on its own, only leaving a simple punctual intensity tuning to the operator, which makes for a prompt acquisition process.

IV. Experiments

In this section, we will specify the content of the datasets that were gathered thanks to our system and how they were used for experimental purposes. The objective of the tests was to assess the performances of CNN models when the parameters of physical light sources varied during the acquisition of the population of their training/validation sets. From here onward, since each couple of training and validation collections contain similar data, to facilitate the reading, they will be referred to as a single entity: "training datasets".

A. Experimental Conditions

The variables of our experiments were as follows:

- **Independent variables**: The effects of different illumination settings, such as their brightness and their orientation, were observed.

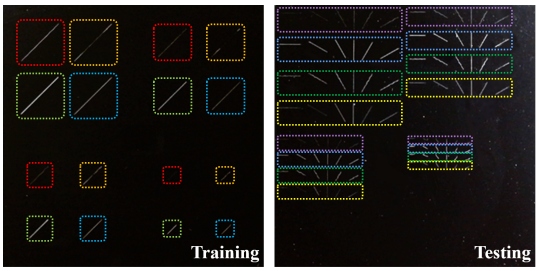

Fig. 5. Samples of training and testing pieces with their lines distributions

- **Dependent variables**: The detection abilities of the different models were measured thanks to the number of True Positives (TP), False Negatives (FN) and False Positives (FP) they output when handling testing data. We also evaluated them through their precision and recall, which are metrics derived from TP, FN and FP.
- **Control variables**: Throughout our experiments, the sizes of the training and validation datasets have been kept at respectively 40 and 5 images. Every model has been trained for 3000 iterations in Darknet, a neural network framework used to train YOLO [16] models. Furthermore, every model has been tested on the same set of 24 pictures to ensure a consistent and reliable baseline to confront the results.

B. Dataset Compositions

We formed 5 datasets based on photographs of six different plastic parts created by cutting sheets of acrylic resin into 100 by 100mm squares. Lines have been etched onto the surfaces of these pieces with a laser engraver, so that they mimic scratches that can usually be found on items on a production line, for example. Every piece comes with its own pattern of lines, carved with determined sets of laser parameters (power, speed and repetition), repeated in four different areas but with decreasing lengths. The lines within the training pieces were 20mm, 15mm, 10mm and 5mm long, while that of the testing pieces were 7mm, 5.25mm, 3.50mm and 1.75mm long. The aim was to obtain synthetic flaws that had different aspects. A visual explanation of the line distributions can be found in Fig. 5, in which features highlighted with a common color were traced with the same laser configurations.

The pictures have all been collected by our semi-automatic visual inspection system. In order to minimize the risks of getting biased datasets, the pictures have been taken at night time when no external source could interfere with the light exposure of the pieces intended by the system.

1) Training data's characteristics: Four pieces comprising 16 scratches each, listed in Fig. 6, were used to take 784 photographs. This amount of data is due to the number of combinations possible with:

- 4 pieces whose scratches are tilted: 0°, 45°, 90°, 135°
- 7 light intensities: 0.45W, 0.9W, 1.35W, 1.8W, 2.25W, 2.7W, 3.15W
- 28 light positions: 0°, 10°, 30°, ..., 250°, 260°, 270°

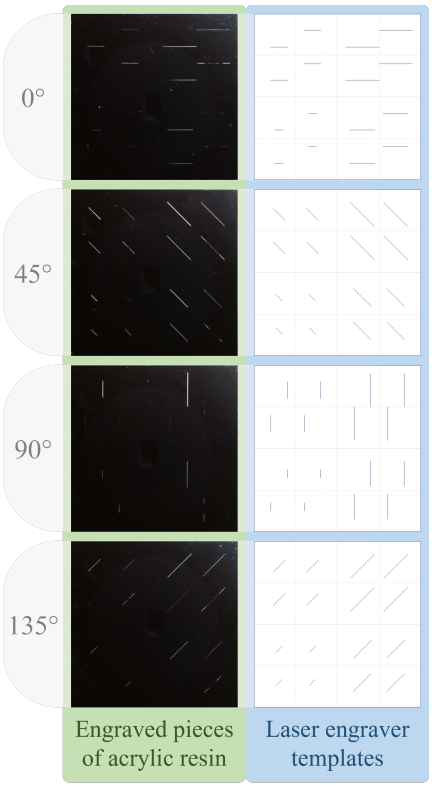

Fig. 6. Base pieces used for training, with the tilt value of their scratches

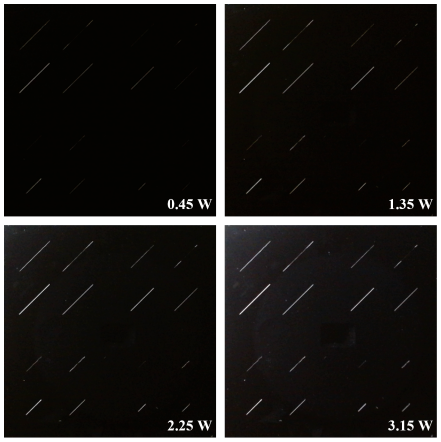

Fig. 7. Training pieces with different light intensities

Fig. 7 presents one of the plastic parts under different illuminations, while Fig. 8 shows the same sample brightened from distinct angles.

Before being incorporated into any training datasets, these data had to be annotated. Since some of them had scratches difficult to see clearly, especially those taken in low brightness, it was decided to create an algorithm that created copies of the images and turned their pixel black if their respective luminance values were below a given threshold, and white otherwise. This way, we were able to use the black and white pictures to consistently classify what scratch should be

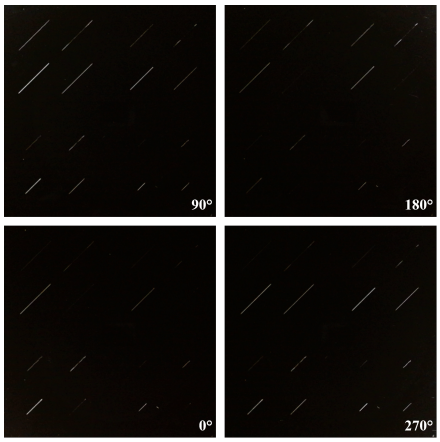

Fig. 8. Training pieces with different light orientations

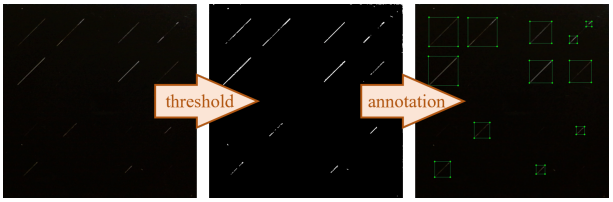

Fig. 9. Scratch annotation based on their luminance value

annotated (white pixels) or which one should be considered "invisible" (black pixels), as shown in Fig. 9. In our case, a value of 30 for the luminance was chosen, as it was below this point a scratch became hardly perceptible to the naked eye.

2) Test data's characteristics: Two pieces comprising 96 scratches each, listed in Fig. 10, were used to take 24 photographs. This amount of data is due to the number of combinations possible with:

- 2 pieces: test_A, test_B
- 4 light intensities (in watts): 0.45, 1.322, 2.25, 3.122
- 3 light positions (in degrees): 0, 100, 200

3) Overview of the datasets: Four training datasets were assembled, whose structures are summarised in Table I. The formations of those batches aimed at getting models trained on images captured in specific light conditions, so as to try to perceive which setting may be the most efficient, when opposing their results. In addition, 1 testing dataset of 24

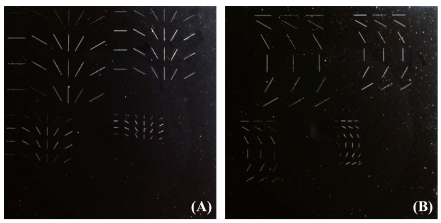

Fig. 10. Testing pieces

TABLE I. OVERVIEW OF THE TRAINING DATASETS' CONTENTS

Dataset name	Light intensities (in watts)	Light orientation (in degrees)	Size
TRAIN_intLow_oriQuad	0.45 / 0.9 / 1.35	0 / 90 / 180 / 270	Train: 40 Valid: 5
TRAIN_intHigh_oriQuad	2.25 / 2.7 / 3.15	0 / 90 / 180 / 270	Train: 40 Valid: 5
TRAIN_intMix_oriQuad	0.45 / 0.9 / 1.35 1.8 2.25 / 2.7 / 3.15	0 / 90 / 180 / 270	Train: 40 Valid: 5
TRAIN_intMix_oriMix	0.45 / 0.9 / 1.35 1.8 2.25 / 2.7 / 3.15	0... *(increments of 10°)* ..270	Train: 40 Valid: 5

TABLE II. OVERVIEW OF THE TESTING SUB-DATASETS, ORGANISED BY LIGHT INTENSITY

Dataset name	Light intensities (in watts)	Light orientation (in degrees)	Size
TEST_intVeryLow	0.45	0 / 100 / 200	6
TEST_intLow	∼1.322	0 / 100 / 200	6
TEST_intHigh	2.25	0 / 100 / 200	6
TEST_intVeryHigh	∼3.122	0 / 100 / 200	6

distinct pictures was gathered to assess the performances of the models. During the computation of the metrics, to emphasise the hypothetical discrepancies between the models (especially the brightness represented in their training images), it has been considered as 7 sub-datasets: 4 organised by light intensity, outlined in Table II, and 3 organised by light orientation, depicted in Table III. The performances of the models are confronted and reviewed in the next section.

C. CNN Trainings

Finally, once the data was collected, annotated and put together, they were used to train models. For our project, the neural network framework named Darknet, and most especially its algorithm/network architecture YOLOv4 [16], has been chosen to assist in the generation of models. This end-to-end learning algorithm is well-known in the field of image processing, its speed and accuracy held it as a reference among CNNs. Every one of the 4 datasets was fed into the algorithm to output its corresponding model, so experiments could be conducted to analyze their respective abilities to find defects within pictures. The training phases were not started from scratch, **yolov4.conv.137** pre-trained weights were chosen as a base. Since the evolution of the average training loss curves indicated that the models had converged by 3000 iterations, a training session stopped when this count was reached.

TABLE III. OVERVIEW OF THE TESTING SUB-DATASETS, ORGANISED BY LIGHT ORIENTATION

Dataset name	Light intensities (in watts)	Light orientation (in degrees)	Size
TEST_ori0	0.45 / ∼1.322 / 2.25 / ∼3.122	0	8
TEST_ori100	0.45 / ∼1.322 / 2.25 / ∼3.122	100	8
TEST_ori200	0.45 / ∼1.322 / 2.25 / ∼3.122	200	8

TABLE IV. METRICS RESULTING FROM THE TESTS CONDUCTED ON EACH MODEL, ON THE ENTIRE TESTING DATASET (24 PICTURES)

Model	TP	FN	FP	Precision	Recall
TRAIN_intLow_oriQuad	597	1707	2	0,997	0,259
TRAIN_intMix_oriQuad	447	1857	9	0,98	0,194
TRAIN_intHigh_oriQuad	502	1802	6	0,988	0,218
TRAIN_intMix_oriMix	644	1660	3	0,995	0,28

V. INTERPRETATION OF THE RESULTS

As a first step to confront the ability of the models to spot irregularities, we had trained *TRAIN_intLow_oriQuad*, *TRAIN_intMix_oriQuad* and *TRAIN_intHigh_oriQuad* so tests could be conducted with the brightness as an independent variable, as their pictures were taken with similar light orientations. Based on the 24 photographs of test plastic pieces, the total number of True Positives (defects found), False Negatives (defects unexposed) and False Positives (regular areas inferred as abnormal) were counted. Then, the precisions and recalls were derived from these TPs, FNs and FPs.

The precision, whose formula is shown in (1), describes the proportion of "positive identifications" that were actually correct.

$$Precision = \frac{TP}{TP + FP} \tag{1}$$

The recall, whose formula is shown in (2), describes the proportion of correct "positive identifications" that were successfully identified.

$$Recall = \frac{TP}{TP + FN} \tag{2}$$

The values of the various metrics are transcribed in the Table IV. If the header of this table is coloured green, it translates a positive metric for classification of a model, while a red-coloured header is for a negative one. When a metric's value is written in green, it means that this value is the best of those recorded; if it is red, it is the worst (it takes into account the positive/negative nature of the metric).

When analysing the measures, it has been noticed that the outcomes of the models were relatively close to one another, nonetheless, *TRAIN_intLow_oriQuad* was the most efficient and *TRAIN_intMix_oriQuad* was the less effective. In order to examine the impact of light orientation, it was attempted to increase the diversity of the less reliable dataset: instead of only including images of pieces that were enlightened from a source positioned at either 0°, 90°, 180° or 270°, we decided to pick from 28 angles. In this way, *TRAIN_intMix_oriMix* has been constituted, as *TRAIN_intMix_oriQuad* was the most lacking. The TPs, FNs, FPs, precision and recall obtained are also reported in Table IV. From this new array of values, it has been noted that *TRAIN_intMix_oriMix* not only outperformed *TRAIN_intMix_oriQuad*, it also seemed to slightly surpass the results of the other models.

Though there were differences between the models, they were rather limited and based on a single set of data, which did not reflect their effectiveness at handling data according to their attributes. It was conjectured that it would be pertinent to underline the capabilities of the models at processing images

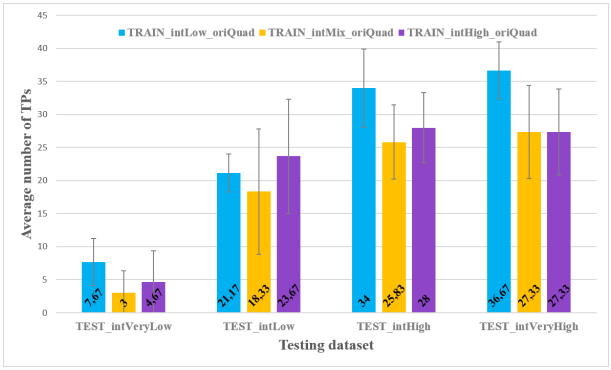

Fig. 11. Comparisons of the average number of TPs on the test subsets arranged by brigthness, grouped by light intensities within their training data

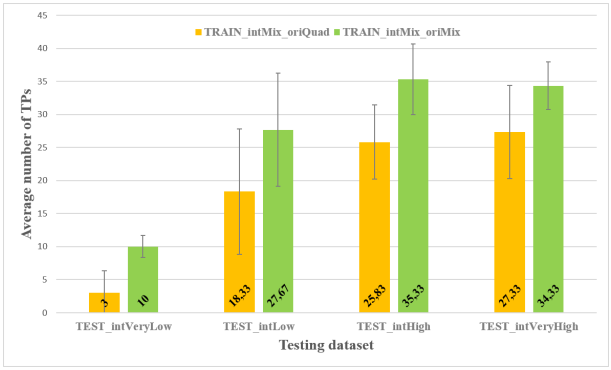

Fig. 12. Comparisons of the average numbers of TPs on the test subsets arranged by brigthness, grouped by light orientations within their training data

taken in specific light conditions. Thus, the former test dataset was split into several subgroups consisting of images sorted either by intensity (Table II) or by orientation (Table III). For these new comparisons, it was decided to restrict the metrics to the average of True Positives: within the scope of our experiments, the significance of this measure makes it sufficient to gain a decent appreciation of the quality of a neural network's training.

The new arrangements of data corroborated the previous experience. Indeed, for the sets containing photographs classified by their luminosity, *TRAIN_intLow_oriQuad* was more efficient than *TRAIN_intHigh_oriQuad* and even more than *TRAIN_intMix_oriQuad*, as shown in Fig. 11. *TRAIN_intMix_oriMix* obtained more satisfactory results than its counterpart trained with four different light orientations, yet again, as visible in Fig. 12. The reason why the models struggled to detect scratches in the case of very low light intensities is that even though most of them were barely discernible, they were annotated as ground truths within the test data, which led to a lot of False Negatives (and, conversely, hardly any True Positives) during the tests that included them. The purpose of these data was to challenge the models on confounding specimens.

Furthermore, similar observations were made for the sets containing photographs classified by the angular positions of

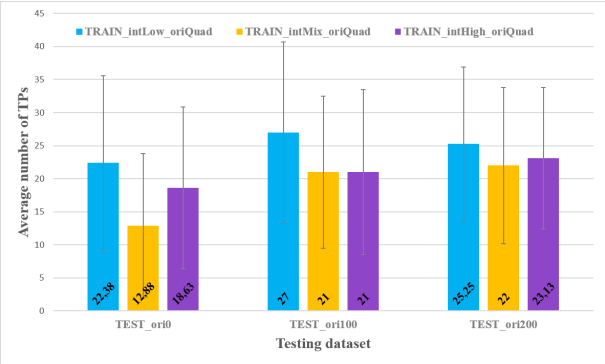

Fig. 13. Comparisons of the average number of TPs on the test subsets arranged by orientation, grouped by light intensities within their training data

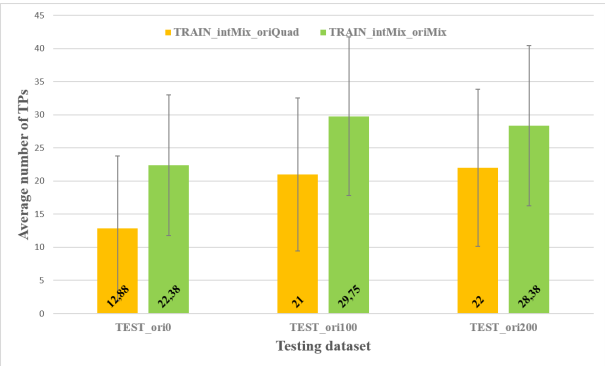

Fig. 14. Comparisons of the average numbers of TPs on the test subsets arranged by orientation, grouped by light orientations within their training data

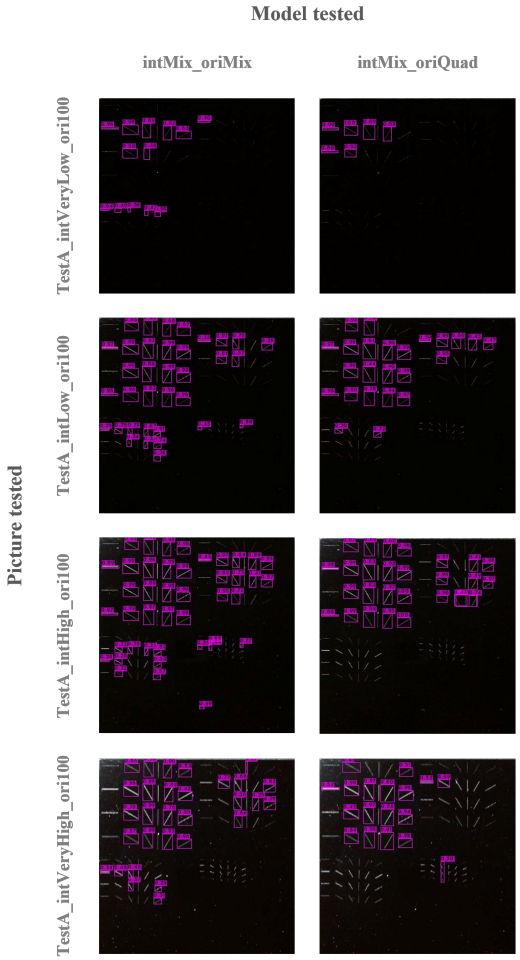

Fig. 15. A few inferences produced by *intMix_oriQuad* and *intMix_oriMix*

their light source, as shown in Fig. 13 and Fig. 14.

A few examples of images with their corresponding inferences, output by *TRAIN_intMix_oriQuad* and *TRAIN_intMix_oriMix*, are displayed in Fig. 15.

VI. Conclusion

Our study involved the conception of an automatic system that facilitates the acquisition of pictures intended for training/testing datasets, while adding physical data augmentation in the form of variations of light brightness and orientation.

This device was used to gather our own images, which were assembled into datasets that were fed to CNNs. These neural networks generated multiple models, each trained on photographs taken under specific lighting conditions. These models were then opposed to one another to observe if the parameters of their illumination had some effects on their ability to handle new data.

It has been noted that both the independent variables (light intensity/light position) that were experimented on influenced the outcome of the training processes. For instance, a model based on data whose illumination varied, but was limited to four positions, obtained worse results than another model based on data whose illumination also fluctuated, but had a broader array of angular positions.

Since the pieces involved in this research have been created to suit a specific scenario: scratches detection on plastic parts, our "hardware data augmentation" technique might only be relevant to this test case. For that reason, future work reiterating the experiments, with larger and more intricate datasets, is expected to confirm (or disprove) the conclusions drawn from the current study. Additionally, the implementation of an automatic light intensity controller, along with the incorporation of YOLO into our software, are objectives envisaged to make our system a fully independent computer-assisted visual inspector.

References

[1] T.L. Johnson, "How and why we need to capture tacit knowledge in manufacturing: Case studies of visual inspection", *Applied Ergonomics* Vol.74, pp.1–9, 2019.

[2] V. Hoskere, "Vision-based Structural Inspection using Multiscale Deep Convolutional Neural Networks", *arXiv*, 2018.

[3] S.Y. Wang, "A computer vision based machine learning approach for fatigue crack initiation sites recognition", *Computational Materials Science* Vol.171, 2020.

[4] I. Kuric, "Visual Product Inspection Based on Deep Learning Methods", *Advanced Manufacturing Processes*, pp.148–156, 2020.

[5] A.R. Ajiboye, "Evaluating the effect of dataset size on predictive model using supervised learning technique", *International Journal of Software Engineering & Computer Sciences (IJSECS)*, Vol.1, pp.75–84, 2015.

[6] C. Shorten, "A survey on Image Data Augmentation for Deep Learning", *Journal of Big Data*, 6:60, 2019.

[7] R. Gontijo-Lopes, "Tradeoffs in Data Augmentation: An Empirical Study", *International Conference on Learning Representations*, 2021.

[8] L. Taylor, "Improving Deep Learning Using Generic Data Augmentation", *arXiv*, 2017.

[9] I. Kandel, "Brightness as an Augmentation Technique for Image Classification", *Emerging Science Journal*, Vol.6 - No.4, pp.881–892, 2022.

[10] D. Crispell, "Dataset Augmentation for Pose and Lighting Invariant Face Recognition", *arXiv*, 2017.

[11] D. Takaki, "Objective scratch and dig measurements", *Savvy Optics Corp.*, 2012.

[12] G. Fu, "A deep-learning-based approach for fast and robust steel surface defects classification", *Optics and Lasers in Engineering*, Volume.121, pp.397–405, 2019.

[13] L. Song, "Weak Micro-Scratch Detection Based on Deep Convolutional Neural Network", *IEEE Access*, vol.7, pp.27547–27554, 2019.

[14] L. Yang, "Steel Plate Surface Defect Detection Based on Dataset Enhancement and Lightweight Convolution Neural Network", *Machines (MDPI)* 10, 523, 2022.

[15] C.J. Barr, "Influence of surface texturing on scratch/mar visibility for polymeric materials: a review", *Journal of Materials Science*, pp.1221–1234, 2017.

[16] J. Redmon, "You Only Look Once: Unified, Real-Time Object Detection", *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.779–788, 2016.