INCREASE OF THE AUTOMOTIVE POWER TRANSISTOR MODULES MANUFACTURE RELIABILITY USING AI DETECTING SYSTEM FOR SOLDERING SPLASHES

Dušan Koniar[1], Peter Klčo[1], Libor Hargaš[1], Marek Chnápko[2], Katarína Pocisková Dimová[2], Vladimír Kindl[3,*]

[1]Department of Mechatronics and Electronics, Faculty of Electrical Engineering and Information Technologies, University of Zilina, Zilina, Slovakia
[2]Semikron, s.r.o., Vrbove, Slovakia
[3]Department of Power Electronics and Machines, Faculty of Electrical Engineering, University of West Bohemia, Pilsen, Czech Republic

*E-mail of corresponding author: vkindl@fel.zcu.cz

Dušan Koniar 0000-0003-3029-3193, Peter Klčo 0000-0001-5386-2243,
Libor Hargaš 0000-0001-8716-0944, Vladimír Kindl 0000-0003-2378-8059

Resume
This study discusses the automated visual inspection of electronic boards used in the mass production of power electronics equipment. Soldering splashes, produced during the necessary manufacturing stages, can affect the hybrid power semiconductor modules' quality, specifications and lifespan. Splashes from soldering may appear in some electronic boards areas when they may cause decreasing of quality of these boards and need to be removed. Image analysis algorithms are used to search for such areas. The automated inspection is based on neural network YOLO. The images of electronic boards, acquired by authors in SEMIKRON Slovakia company, were used as training, testing and validation dataset for the YOLO network. The custom recording system was designed to acquire those images. The implementation of this method may increase the object search success and then thereby increase the overall quality of module production.

Article info
Received 5 December 2022
Accepted 25 January 2023
Online 1 March 2023

Keywords:
transistor
module
solder
splash

Available online: https://doi.org/10.26552/com.C.2023.032

ISSN 1335-4205 (print version)
ISSN 2585-7878 (online version)

1 Introduction

The first IGBT based standard module appeared in 1989. Today the construction is still closely related to the original set-up. One or more ceramic plates DCBs (Direct Copper Bond) are soldered onto a copper base plate, which also serves as the mounting surface to the heatsink. They consist of a thin insulating substrate made of Al_2O_3, Si_3N_4 or AlN and copper with variable thickness coated on both sides.

Soldered to the top of the DCBs are the bottom sides of the chips (IGBTs: collectors, diodes: usually cathodes). The chip tops sides (IGBTs: gate and emitter, diodes: anode) are contacted by parallel Al or Cu bond wires to corresponding pads on the DCB substrate. Either soldered, or ultrasonic welded, directly onto these pads, are the module terminals. There are more soldered, ultrasonic welded or bonded connections to the terminals mechanically fixed in the module housing (Figure 1).

The continuous development of the construction and connection technology of power modules is driven by the cost measures in production and processing of the modules on the one hand and ever-increasing demands on the module properties on the other.

This involves the best possible dissipation of the heat loss in the semiconductors to the cooling, both statically, as well as for the short-term loads. Power modules can be utilized very effectively with efficient cooling, which is a “must” for the high power and lowest possible semiconductor cost. Thus, the permissible temperatures of the IGBT and diode chips are increased from one chip generation to the next. However, high chip operating temperatures and efficient heat dissipation tighten the requirements for temperature and power cycling capability of the module designs. This requires

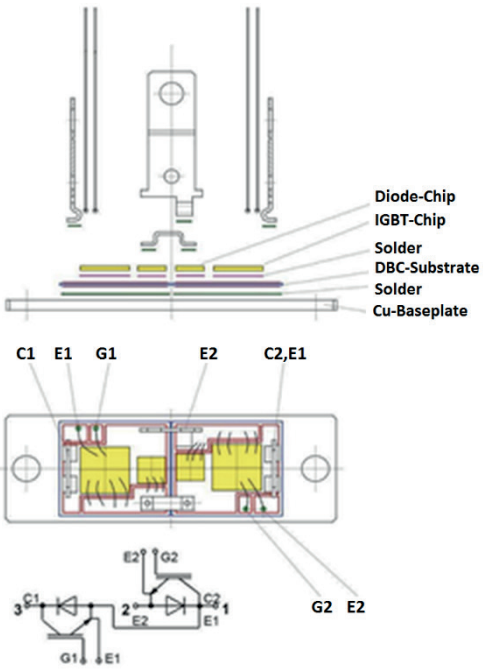


Figure 1 *Basic concept of soldered module*

continuous development of the construction and connection technologies.

One of the most important processes in power modules production is soldering (Figure 2). Soldering is currently one of the basic methods of connecting circuit elements. When soldering power modules, it is necessary to ensure a perfect connection of individual components with the baseplate. In practice, it sometimes happens that during the soldering of individual elements there is an imperfect connection of components or splashes are formed. Detection of soldering splashes is important for increasing the quality and reliability of power electronic modules and it is important from the economical point of view, as well: most of detected soldering splashes can be removed and electronic board of expensive module can be used again.

Industrial manufacturing processes are supervised by the visual inspection systems, to identify possible defects. To increase reliability, reduce the number of defects (higher identification rate) and decrease the time required for this procedure, progressive detection algorithms and visual systems are being developed within last years. Last decade, the computer vision tasks for specific objects detection are based on neural networks (especially convolutional neural networks - CNN). There are some works comparing the modern CNNs and conventional classifiers (such as Support Vector Machine - SVM) used in algorithms for searching specific objects in images. Ren et al. [1] concluded that selected CNN obtained better results for identification of given object as method based on SVM. Multilayer feedforward neural networks can be computationally demanding due to high number of weight parameters. The neuronal representation of spatial information can be another bottleneck for flattened layers [2]. Due to significant expansion and massive implementation of CNNs to many image identification tasks, it exists several CNN architectures and their derivatives. The YOLO architectures, which were used for our application, are discussed in the following. The research shown their applicability to inspect the electronic boards (PCBs) for different types of defects. The YOLO model (You Only Look Once) belongs to deep neural networks designed by Joseph Redmon as one step process classification algorithm. The input image is divided into a grid SxS. The YOLO network can detect multiple bounding boxes in single grid cell. In comparison to other deep neural networks, the YOLO's detection speed is relatively high. The YOLO can classify multiple categories simultaneously [3]. The final stage of model classification results in localization of bounding box and type of category identification [4]. In our case there is only one object category - soldering splash. The YOLO classification model is considered as one of the most successful modifications of Convolutional neural networks [5]. Adibhatla et al. proposed deep learning model based on Tiny-YOLOv2 architecture for PCB quality inspection. Their dataset consisted of 11000 images (420x420 pixels) obtained from the automated optical inspection (AOI) machine. The model achieved 98% accuracy for batch size 32. The YOLO model can have lower recall or increased localization errors than the Fast R-CNN network model [2]. The YOLO architecture can be preferred due to tradeoff between detection accuracy and speed. The authors consider template-matching and image subtraction as traditional methods for the PCB defect detection [6-8]. Liao et al. used improved YOLOv4 algorithm for detection of 6

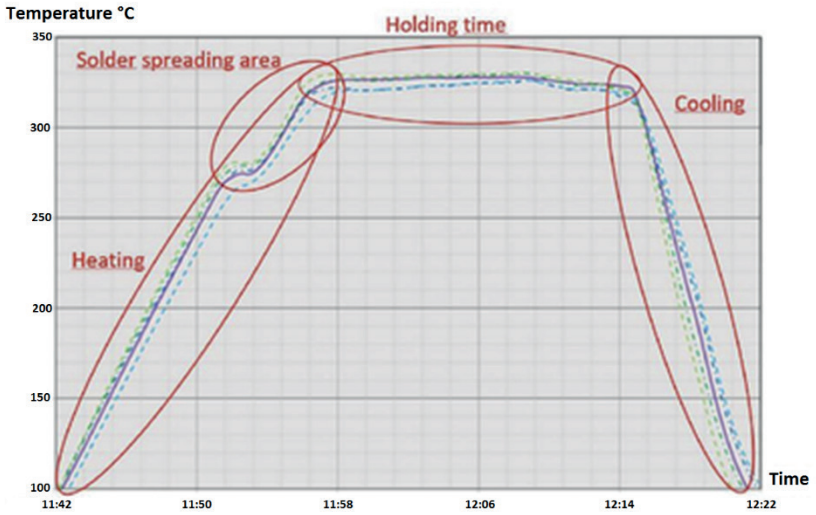


Figure 2 *Timeline of automated soldering procedure*

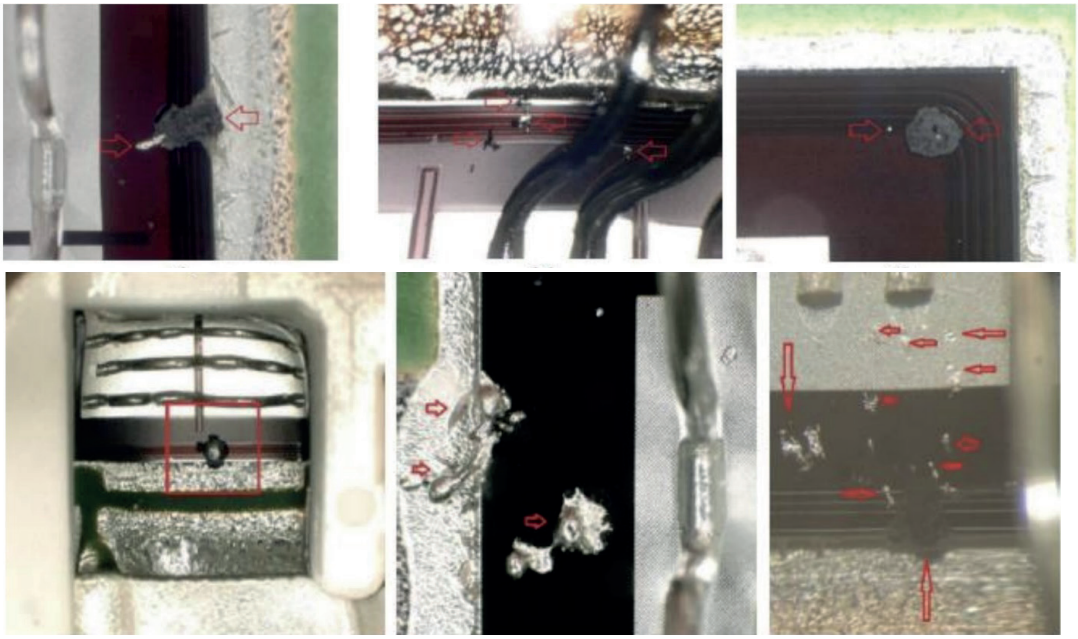


Figure 3 *Typical examples of soldering splashes in inspected images*

different PCB surface defects - scratch, clutter, broken line, hole loss, line repair damage and over oil-filling. The original dataset consisted of 2008 PCB defect images. The data augmentation was performed and number of images increased to 19029. Originally 12-megapixel images were resized to 416 x 416 pixels and labelled with LabelImg program. The neural network model with 39.5 million of weight coefficients achieved high performance with 98.6 % mean average precision score (mAP) [9].

2 Soldering process characteristics

In production soldering process, exist three procedures:

- One-step soldering process: whole power module with chips and terminals (power terminals and signal terminals) is assembled in soldering mask and soldered together by the preform solder.
- Two-step soldering process: power module is assembled in two steps, the 1st step consists of silicone chips soldering by the solder paste, the 2nd step consists of preform soldering from power hybrid to baseplate and terminals (main power terminals and auxiliary/signal terminals).
- The second soldering process: used for power modules with high void ratio under the DCB or chips. This process is introduced for decreasing the scrap ratio after the soldering process and eliminate unwanted voids.

Solder under the chips consists of solder and flux (solder paste). Solder under the DCB is without flux (preform solder used for Semitrans10, SEMiX Press-fit). In two-step soldering process (or second soldering

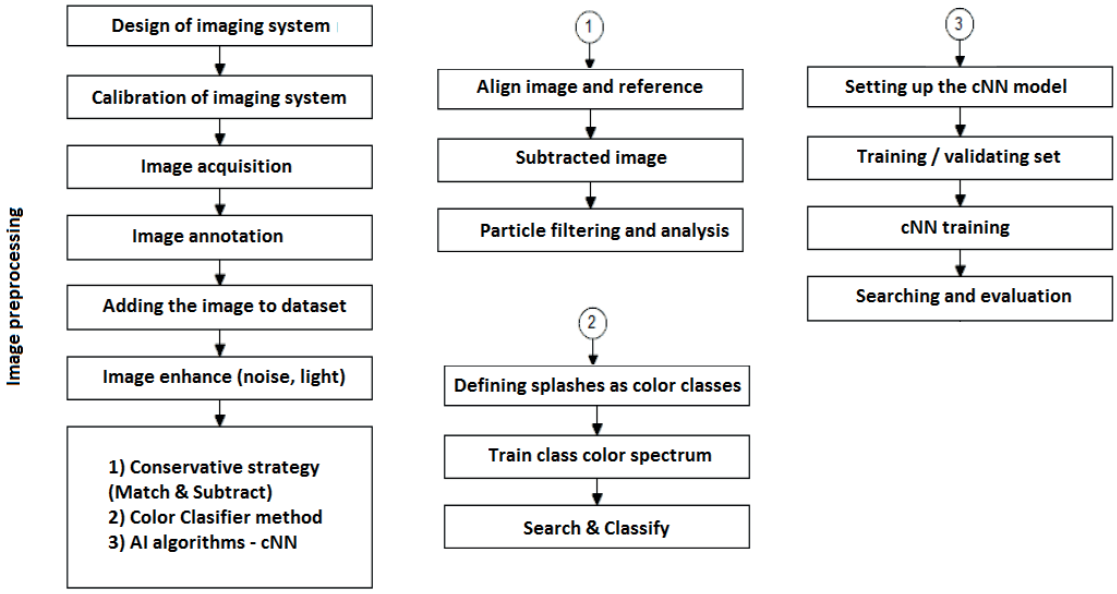


Figure 4 *Image acquisition, preparation of image dataset and 3 image analysis concepts*

process) during the soldering (or re-soldering) of power hybrids (PH) to the baseplate the solder under the chips is melted again. Flux is flowing out from the solder to the chips' edges. If there is some restriction or resistance to the flux movement, it can cause increase of internal mechanical forces and solder splashes. Solder splashes can be found on more than one production line and for different solder alloys and for different soldering equipment (Centrotherm oven or Pink oven) (Figure 3). It can be found for all the power hybrids and chips dimensions.

In production there are restricted areas, directly defined in product drawings (power hybrid drawing, module drawing, etc.), where solder splashes cannot be found. If the solder splashes are in the restricted area it is necessary to clean these areas. Proposed operations for cleaning, actually used and tested, are milling (used in production) or ultrasonic washing (actually tested for various products).

- Solder splashes most often cause problems with micro welding, bonding and ultrasonic welding. Bigger solder splashes can cause micro welding and bonding head damage.
- Additional problem, caused by solder splashes, is when they are located on Data Matric Code (DMX) position. This can cause damage of the DMX code and restrict or wipe out traceability in production process.
- Another effect is for position of solder splashes on power hybrid (PH) where it can cause damage on chip passivation, decreasing of minimal electrical distances, decreasing the module lifetime etc. This causes decreasing of power module reliability and quality.
- Solder splashes located on baseplate will directly influence thermal characteristics of the power module, increase in thermal resistance, as well as increase in thermal paste thickness and cooling in target customer application.

3 Proposal for imaging system and preprocessing of image information

When designing the chain of acquisition and preprocessing of image information, it is important to precisely plan individual steps in advance, so that the final result meets all the requirements. Since we are working with LabVIEW software, the following steps will also be closely related to its use. To obtain a data source of images, it is necessary to prepare the equipment for scanning, which consists of a stand, provision of constant light, auxiliary template reference white color, dark uniform background, defective VPM with containing object of interest, industrial camera from the company Basler type: ace acA4600-10uc together with C125-1218-5M 12 mm lens and processing software, marking area of interest and storing image information in the necessary data form.

The entire creation process is shown in Figure 4. How the system is set up depends on imaging environment and the type of analysis and processing to be performed. The imaging system should produce images of sufficient quality so the required information can be extracted.

3.1 Sensing chain

Sensing chain is processing several steps from sensed information:

- Design of sensing apparatus

Table 1 *Parameters of sensing apparatus - Basler camera*

Parameter	Specification
Video output	USB 3.0
Manufacturer	Basler
Type of sensor	Progressive Scan CMOS
Format of camera sensor	1 / 2.3”
Pixels	4.6 x 3.3
Sensing area (mm)	6.45 x 4.6
Depth of pixel	12 bit
Resolution (megapixels)	15.10

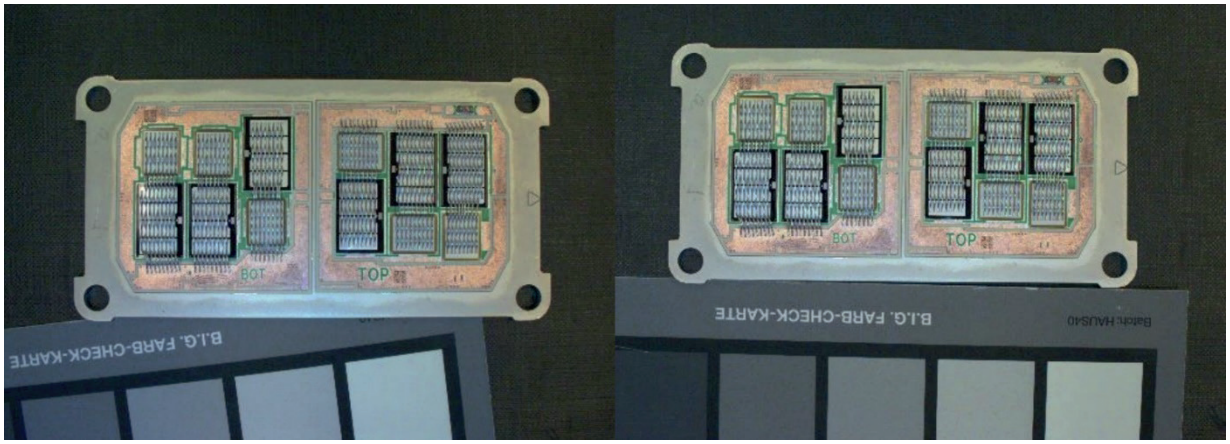


Figure 5 *Comparison of images for 8 mm focal length (left) and 12 mm focal length (right)*

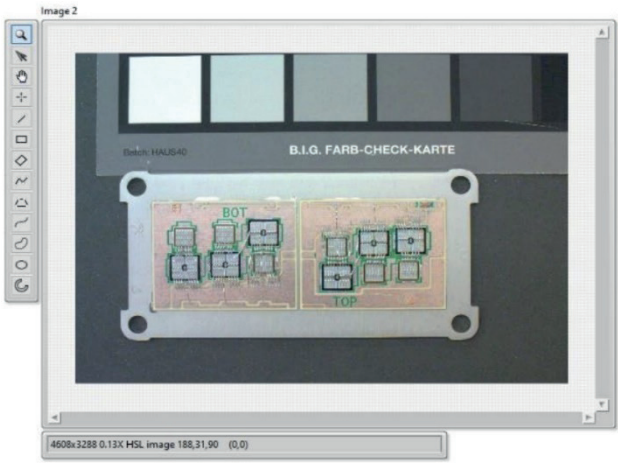


Figure 6 *Preprocessed image after the noise filtration and color balancing (normalization)*

- Pre-processing and storage of sensed information
- Detection and acquisition of position of object of interest
- Camera Basler, type: ace acA4600-10uc with C125-1218-5M 12 mm objective was used as a sensing apparatus. Main parameters are listed in Table 1.

Lenses with a focal length of 8 mm and 12 mm were used in the design of the sensing apparatus. The difference between images with 8 mm focal length and 12 mm focal length is shown in Figure 5.

Pylon Viewer is a software tool, which was used for storage of image information. It has high compatibility with camera and enables automatic modification of the image, what improves white color correction requirements. Stored image is converted to .bmp format, while the object of interest is defined by the technology engineer. The next step is then gaining the coordinates of the objects of interest. Since the deep learning algorithms have been analyzed in this study as well, a set of databases of power module images was developed (approximately 200 recorded modules).

Preprocessing of image information is realized with the use of LabVIEW environment, using NI-IMAQ and Vision Development libraries. The principle of

the sequential loading of images algorithm consists in incrementing the input - the value of the number in the for loop, which is added to the string and thereby creates the whole string path to the new image frame. Images are saved in bmp format with names from 0000 to 0197. The white balance algorithm is based on dividing the image into 3 basic components: red, green and blue, to compare the area of interest and thus our reference white color, the value of which should be 255 points. Reference values of 255 points are divided with the mean value of the given colors. After getting new colors that will be added to the original ones obtained, the image is recomposed using the field-to-image transfer function and their subsequent replacement into the original image (Figure 6).

4 Image database design

To realize the detection algorithms, it is required to modify acquired images by crop and mask tasks, i.e. unnecessary areas, which can influence detection of the objects of interest, are eliminated (Figure 7).

Masking is similar to image straightening on the coordinates of comparison of the extreme point patterns and by calculating their distances, the algorithm can automatically adjust dimensions and placement masking for each frame separately, using the reference percentage value. Once the necessary straightening and masking coordinates are obtained, they can be implemented into a color image, since we have been working in the gray-tone area so far for more accuracy comparing endpoint patterns. The resulting image can be cropped using coordinates and saved to a new file with a new name, making the data source ready for the phase practical implementation of algorithms for detection of objects of interests.

There are several ways how to identify and detect specific objects among which the used algorithms are described later. Algorithms are then utilized for advanced intelligence algorithms for neural networking and predictions of soldering splash detection.

4.1 Global thresholding

A global thresholding technique is one that uses a single threshold value for the whole picture. By correctly setting the colors, red, green and blue, in the form of RGB or hue, saturation and lightness in the form of HSL, the object can be extracted from the background by a simple operation that compares the image values with a threshold value T. Pixels subject and background have gray levels grouped into two dominant modes. One obvious way to extract an object from the background is to choose a threshold T, which separates these modes. Thresholding results in a binary image where pixels with a value intensity of 1 correspond to objects, while pixels with a value of 0 correspond to the background.

As can be seen in Figure 8 it is possible to utilize

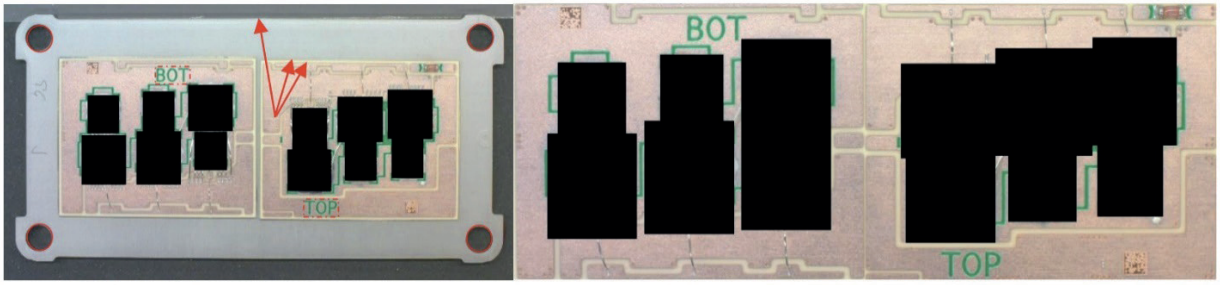


Figure 7 *Comparison of original and modified image with crop and masking algorithms*

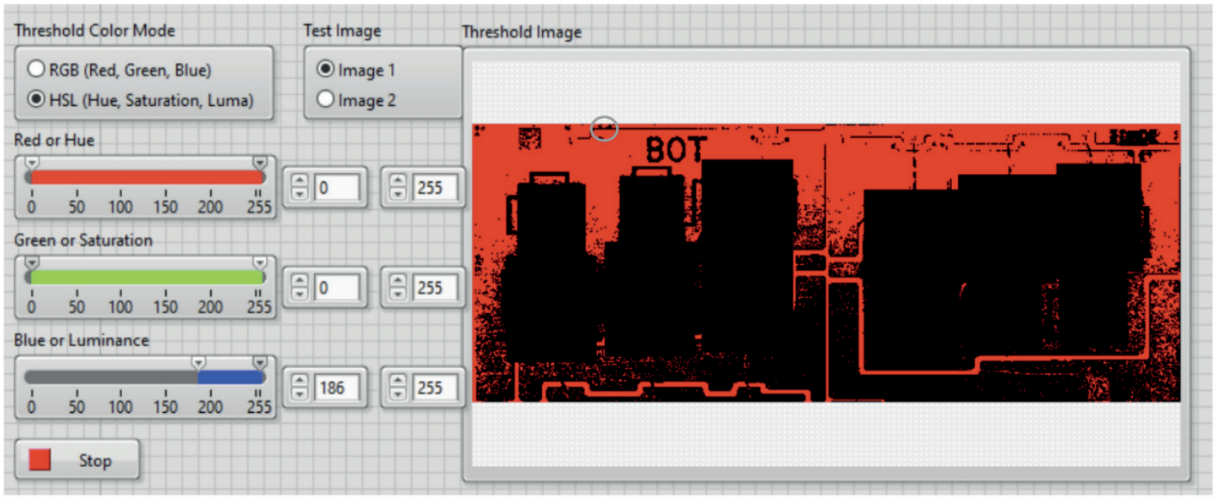
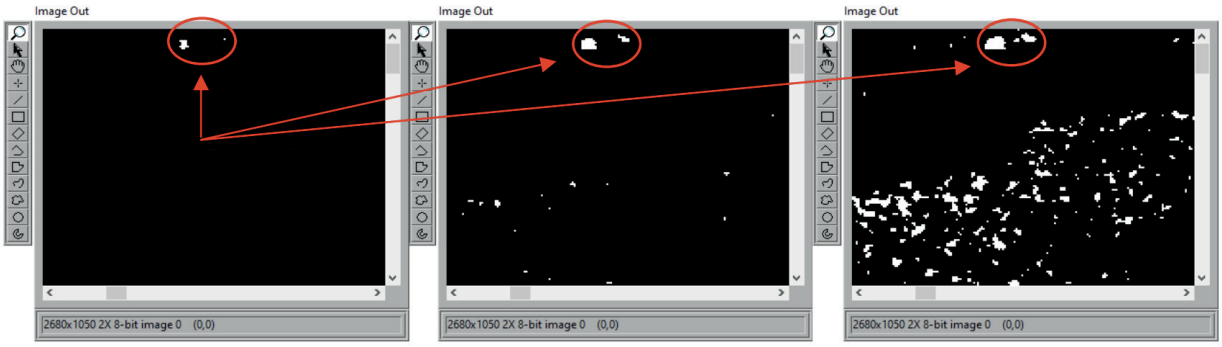


Figure 8 *Example of global thresholding and decomposition into object and background*



a) *b)* *c)*

Figure 9 *Manual thresholding a) 250 threshold b) 240 threshold c) 230 threshold*

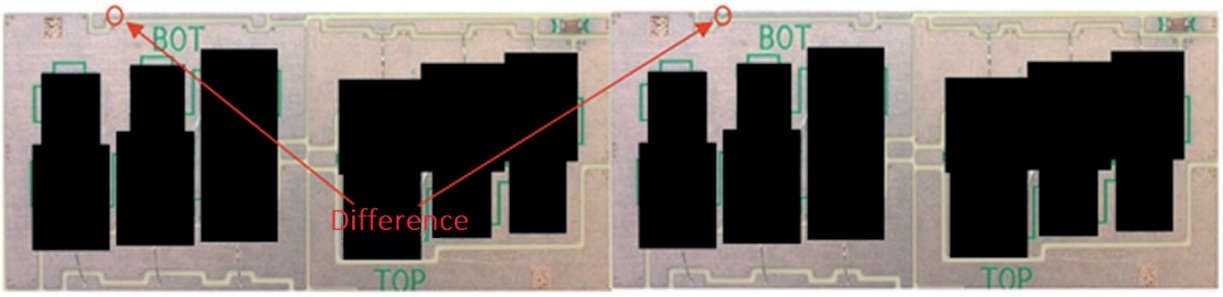


Figure 10 *Comparison of the samples for image difference operation*

the global thresholding to separate the light component from the dark one and thus the solder spatter is in the foreground with a value of 1 and the dark part that surrounds the splash has a value of 0. The disadvantage of the method is that it divides components only in 2 parts, if there is a small difference between the object of interest and the environment, components merge.

4.2 Manual thresholding

Assertion on dependence on the uniformity of lighting is resulted on Figure 9. For clear and relatively large solder joints splashes, it is possible to record their occurrence and thereby separate them from the rest of the image. However if those sprinkles are located in the upper part, where the luminosity is, in our case, the sharpest, it is difficult to separate them from the environment at lower threshold values, what prevents the detection of faulty elements in the image.

When detecting faults on others images, there was a problem not only in uneven lighting but in the variety of colors, size and color difference of the solder splash, as well, whether as a result of the melt or splash location on the power semiconductor module.

4.3 Image difference operation

It works on a pixel-by-pixel basis of reading two images. The important part is so that the images are symmetrical “pixel by pixel” to ensure the expression

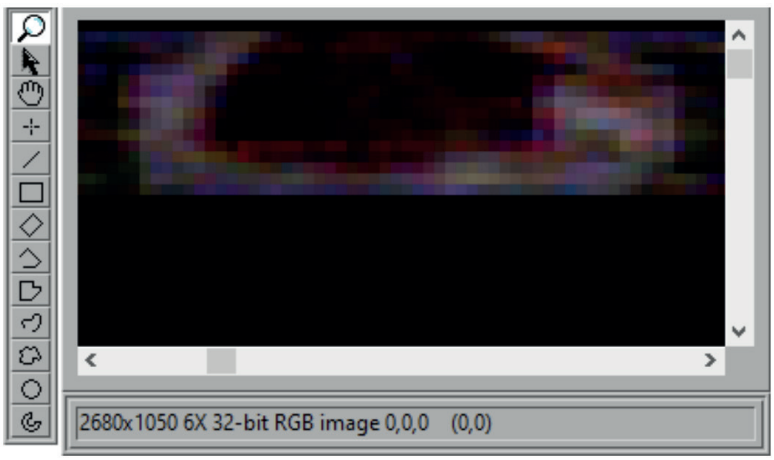


Figure 11 *Result of image difference operation*

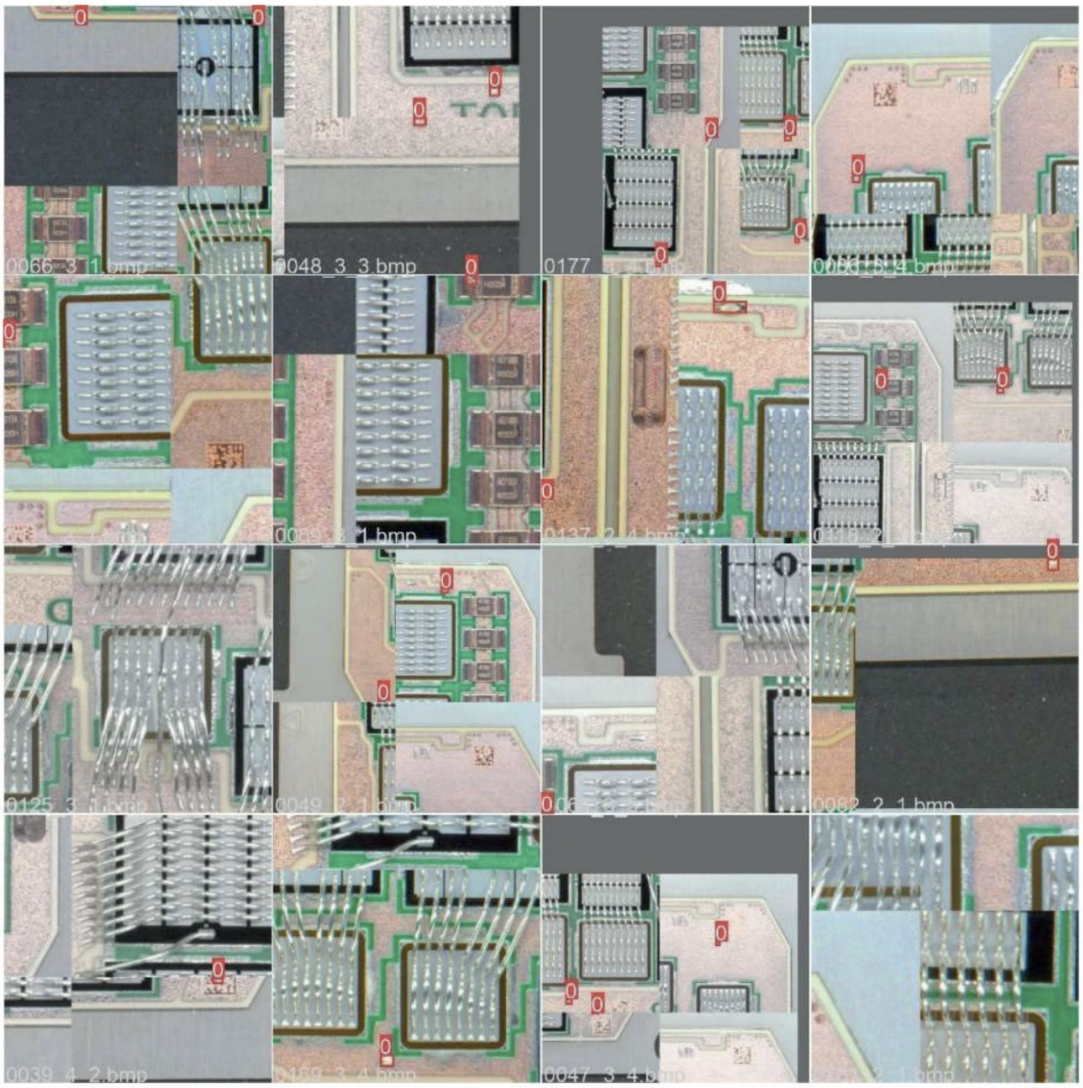


Figure 12 *Example of dataset of images from neural network training*

Table 2 *Model parameters*

Method	Setting
Initial learning rate	0.0001
Optimizer	Adam
Batch	16
Epochs	1000
Image HSV value augmentation	0.0
Patience	300
Image resolution	640

of object difference and to have the same dimensions. The basis is the correct preparation of the data source, which was secured using the multiple coordinates of edge detectors, by coincidence selected patterns, rotation of the image and adding masking. Initially, modified image, with separated solder splash, was used together with the original image (Figure 10). This created an identical match with the difference of the currently searched defect.

Figure 11 shows an enlarged single bright point, which is the soldering splash, whose data information remained after reading the frames. When trying to read two different modules in the picture, it is possible to see the deviations that are caused by different installation of elements on power module, surface changes, material changes and moving the x and y axes. Despite the imperfect match of the two images, it is possible to recognize the object of interest on the image.

5 Neural network training and proposal for AI detection algorithms

Within the proposed system, we decided to use the last versions of the YOLO model, i.e. YOLO v5. It is the most advanced version of this classification algorithm and Python software support is useful in scientific research area. The YOLOv5 medium model was chosen as compromise between the detection speed and accuracy. After the neural network model selection, the training process must be provided. Training process requires a huge dataset of samples. There are many official datasets for various image classification problems, but for detecting the soldering splashes we had to create our own dataset.

After the neural network model selection, the training process must be provided. Training process requires a huge dataset of samples. There are many official datasets for various image classification problems, but for detecting soldering splashes we had to create our own dataset (Figure 12).

As we declared in the previous section, each image in the acquired dataset was annotated by SEMIKRON Company specialist with LabelImg tool dedicated for YOLO. We defined just one class - soldering splash. The original resolution of PCB images was 4608x3288 pixels. The complete dataset actually consists of 231 images and there are 350 soldering splash objects annotated.

Nvidia GTX 1070 with 8 GB of memory was used as a GPU to speed up the training of YOLO model parameters. It was necessary to install the PyTorch, CUDA and OpenCV software libraries. The YOLOv5 neural network model could not be trained on the original image resolution due to the high-resolution quality of the input images and the huge video memory requirements [10]. In order to create 640x640 pixels images, the PCB images were tiled using the YOLO image tiling algorithm [11]. The original resolution of the object was preferred since soldering splashes are rather small-scale objects. The training, validation and testing sets were randomly assigned the following ratios: 70 % - 20 % - 10 % of the input dataset.

The YOLOv5 model was initially trained using the initialization parameters' default values. In a computer simulation experiment, these parameters were modified to determine the ideal starting point for the neural network training. The values of the YOLOv5 model's configurable parameters that produced the best classification performance are shown in Table 2. This configuration was used ten times to replicate the model learning.

Precision was 92.7 %, recall was 87.9 % and the

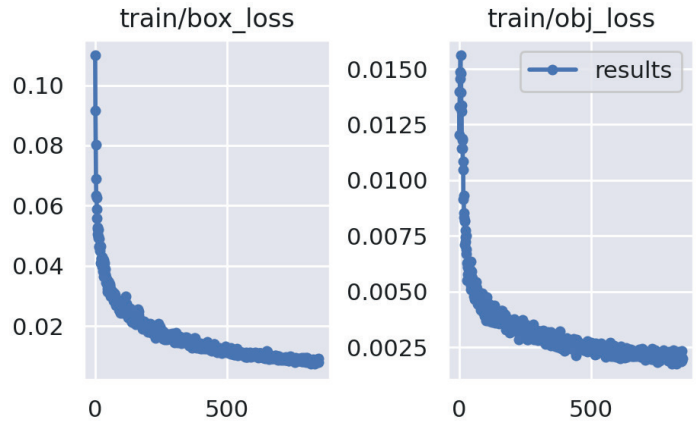


Figure 13 *Plots of objectness and box loss over the training epochs for training set*

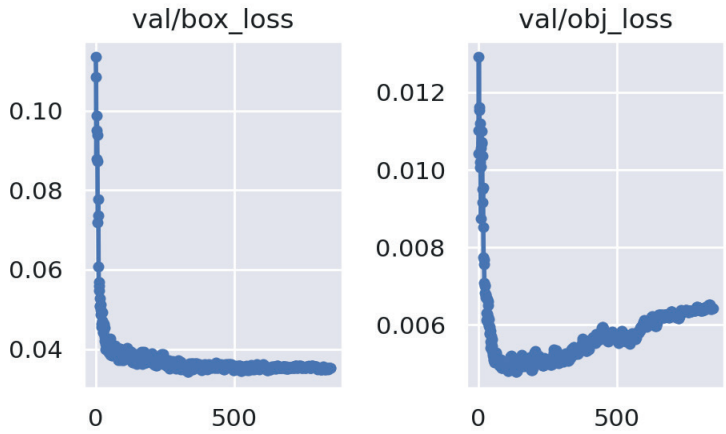


Figure 14 *Plots of objectness and box loss over the training epochs for validation set*

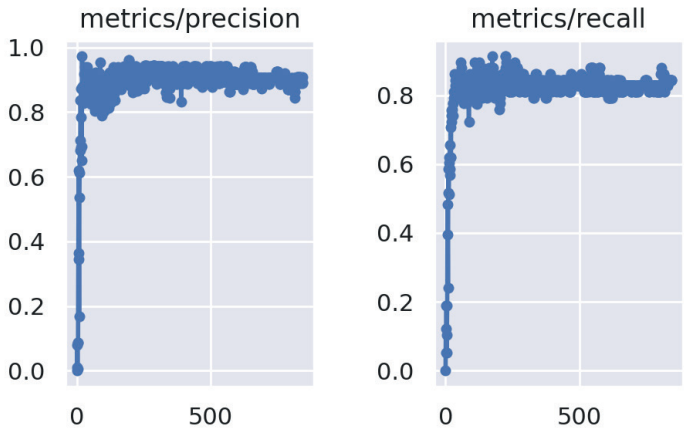


Figure 15 *Plots of precision and recall over the training epochs for training set*

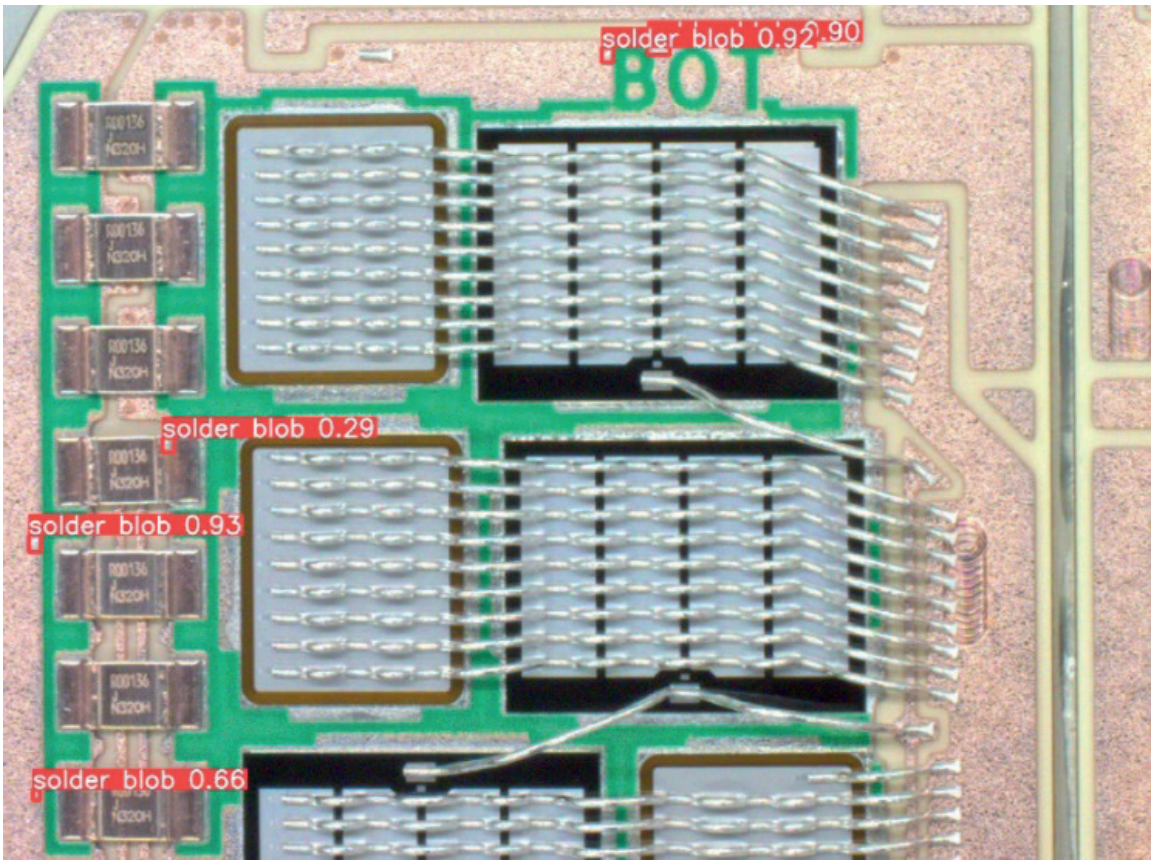


Figure 16 *Example of detected soldering splashes with YOLOv5 model, the solder splashes detected by the computer are marked by red rectangles, numbers indicate confidence score*

mean average precision score (mAP) was 90.6% with these model parameters. Model correctness is described by the mean average precision parameter; a higher value denotes a more accurate neural network model. The following formulas were used to calculate the metrics:

$$Presicion = \frac{True\ Positive}{True\ Positive + False\ positive}, \tag{1}$$

$$Recal = \frac{True\ Positive}{True\ Positive + False\ Negative}. \tag{2}$$

Precision score is represented by Equation (1), where the true positives refer to soldering splashes that were accurately detected, but false positives relate to image areas that were incorrectly identified as soldering splashes. The recall score is represented by Equation (2), where false negatives are related to undetected solder splash items and the remaining terms are the same as those in Equation (1).

Figures 13, 14 and 15 display the pertinent metrics. The training process revealed a declining trend for the training set's normalized box loss score over the epochs (Figure 13). The box loss function shows how well the predicted bounding box encloses an object and

how precisely the algorithm can find the object's center (soldering splash). A measure of the likelihood that an object exists in a suggested region of interest is called the objectness loss (obj loss) (ROI) [12].

The normalized box loss score trend over the training epochs is shown in Figure 14. The box loss also decreased in the training and validation sets over the training phase. This suggests that the YOLOv5 model has strong generalizability potential in terms of pinpointing soldering splash centers. After 100 epochs, the objectness loss in the validation set showed an upward trend, therefore the training was cut short.

Throughout the training, the precision and recall scores were steadily dropping. The recall score could rise if there are more photographs in the collection. The precision and recall score are shown to grow over the training epochs in Figure 15. The trained model was very accurate at identifying the soldering blobs.

Even soldering droplets with small areas were successfully recognized by the specially trained YOLOv5 model (Figure 16). These outcomes were related to the direct application of the YOLOv5 medium model. No image manipulations or alterations were used. The input photos contain several items with similar color and texture to soldering splashes, which can impair classification ability.

6 Discussion and conclusion

Although the YOLOv5 algorithm is not a novel technique, in this article we show the use of machine learning-based object detection in an original dataset of particular PCBs. The study demonstrates that the YOLOv5 algorithm was successfully employed for item recognition, including people, cars, animals and more. This algorithm's use for tiny items, such as soldering droplets, is innovative in the field.

Due to the scarcity of official photos of that kind, we first constructed an annotated dataset of PCBs (including 4 categories of boards).

The trained neural network model achieved precision of 92.7%. Considering the relatively tiny training set and small region of detected objects, these results can be deemed satisfactory. Additionally, several items in the PCB photos, particularly wires and connections, had color or texture characteristics that were similar to soldering splashes.

Using the alternative classification model RetinaNet might improve classification performance. A larger dataset and improved spatial resolution for the PCB photos both have the potential to improve the classification accuracy. Further study will be conducted on this.

Acknowledgment

The results presented in this paper were supported by grant No. APVV 20-0500 - Research of methodologies to increase the quality and lifetime of hybrid power semiconductor modules. This research has also been supported by the Ministry of Education, Youth and Sports of the Czech Republic under the project OP VVV Electrical Engineering Technologies with High-Level of Embedded Intelligence CZ.02.1.01/0.0/0.0/18_069/0009 855. Authors want to thank all SEMIKRON Slovakia company staff for cooperation.

Conflicts of interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] REN, Z., FANG, F., YAN, N., WU, Y. State of the art in defect detection based on machine vision. *International Journal of Precision Engineering and Manufacturing-Green Technology* [online]. 2022, **9**, p. 661-691. ISSN 2288-6206, eISSN 2198-0810. Available from: https://doi.org/10.1007/s40684-021-00343-6

[2] ADIBHATLA, V. A., CHIH, H. C., HSU, C. C., CHENG, J., ABBOD, M. F., SHIEH, J. S. Defect detection in printed circuit boards using you-only-look-once convolutional neural networks. *Electronics* [online]. 2020, **9**(9), 1547. eISSN 2079-9292. Available from: https://doi.org/10.3390/electronics9091547

[3] REDMON, J., DIVVALA, S., GIRSHICK, R., FARHADI, A. You only look once: unified, real-time object detection. In: IEEE Conference on Computer Vision and Pattern Recognition: proceedings [online]. IEEE. 2016. eISSN 1063-6919, p. 779-788. Available from: https://doi.org/10.1109/CVPR.2016.91

[4] HUANG, Y. Q., ZHENG, J. C., SUN, S. D., YANG, C. F., LIU, J. Optimized YOLOv3 algorithm and its application in traffic flow detections. *Applied Sciences* [online]. 2020, **10**(9), 3079. eISSN 2076-3417. Available from: https://doi.org/10.3390/app10093079

[5] HUANG, R., GU, J., SUN, X., HOU, Y., UDDIN, S. A rapid recognition method for electronic components based on the improved YOLO-V3 network. *Electronics* [online]. 2019, **8**(8), 825. eISSN 2079-9292. Available from: https://doi.org/10.3390/electronics8080825

[6] MALGE, P. S. PCB defect detection, classification and localization using mathematical morphology and image processing tools. *International Journal of Computer Applications* [online]. 2014, **87**, p. 40-45. ISSN 0975-8887. Available from: https://doi.org/10.5120/15240-3782

[7] TAKADA, Y., SHIINA, T., USAMI, H., IWAHORI, Y. Defect detection and classification of electronic circuit boards using keypoint extraction and CNN features. In: 9th International Conferences on Pervasive Patterns and Applications Defect PATTERNS 2017: proceedings. 2017. ISBN 978-1-61208-534-0, p. 113-116.

[8] ANITHA, D. B., MAHESH, R. A survey on defect detection in bare PCB and assembled PCB using image processing techniques. In: 2017 International Conference on Wireless Communications, Signal Processing and Networking WiSPNET: proceedings [online]. IEEE. 2017. ISBN 978-1-5090-4443-6, eISBN 978-1-5090-4442-9, p. 39-43. Available from: https://doi.org/10.1109/WiSPNET.2017.8299715

[9] LIAO, X., LV, S., LI, D., LUO, Y., ZHU, Z., JIANG, C. YOLOv4-MN3 for PCB surface defect detection. *Applied Sciences* [online]. 2021, **11**(24), 11701. eISSN 2076-3417. Available from: https://doi.org/10.3390/app112411701

[10] MASTERS, D., LUSCHI, C. Revisiting small batch training for deep neural networks. *ArXiv* [online]. 2018, 1804.07612. eISSN 2331-8422. Available from: https://doi.org/10.48550/arXiv.1804.07612

[11] YOLO dataset tiling script [online]. Available from: https://github.com/slanj/yolo-tiling

[12] Manish Chablani: YOLO - you only look once, real time object detection explained [online]. Available from: https://towardsdatascience.com/yolo-you-only-look-once-real-time-object-detection-explained-492dc9230006