International Journal of Current Innovations in Interdisciplinary Scientific Studies

Volume 5, Issue 1, (2021) 01-15 www.ij-ciss.eu

The performance comparison of optical character recognition and object detection methods on rubber products

Volkan Özdemir *, Kocaeli University, Entrepreneurship Manager - Brisa Bridgestone Sabanci, İstanbul, Turkey

Anil Çalişkan, Bridgestane Sabanci Tire Manufacturing and Trading Inc, Turkey

Murat Özgen, Bridgestone Sabanci Tire Manufacturing and Trading Inc, Trukey

Enes Baytürk, Kocaeli University, Turkey

Sinan Özen, Kocaeli University, Turkey

Suggested Citation:
Ozdemir, V., Çalişkan, A., Özgen, M., Baytürk, E. & Özen, S. (2021). The performance comparison of optical character recognition and object detection methods on rubber products. *International Journal of Current Innovations in Interdisciplinary Scientific Studies. 5*(1), 01-15. Available from: www.ij-ciss.eu

Received from December 21, 2021; revised from February 15, 2021; accepted from May 05, 2021.
Selection and peer-review under responsibility of Dr. Vasfi Tuğun, Kyrenia University, Girne, North Cyprus

Abstract

In the manufacturing sector, there are dedicated codes or characters on finished and semi-finished products that are specifically defined by manufacturers. These codes are composed of characters that have special meanings in themselves and they are usually made up of different combinations of numbers and letters. This study aims to detect the engraved letter and number combinations from the surfaces of cured rubber specimens and rubber products. The data set has been prepared from products and specimens in the production line where the controlling system will be set. The study covers two different approaches which are OCR and object detection. To decide which method satisfies the requirements, detailed analyzes were made along with the detection performances and their robustness against scene changes that may occur due to the nature of the problem. Based on the findings, the performance of OCR methods has not been found as satisfactory.

Keywords: Deep Learning; Low Contrast; Object Detection; OCR; Rubber Sample.

* ***** ADDRESS FOR CORRESPONDENCE: Volkan Ozdemir, Entrepreneurship Manager - Brisa Bridgestone Sabanci, İstanbul, Turkey
E-mail address: v.ozdemir@brisa.com.tr

1 Introduction

Rapid developments in the world of technology and information have made it easier for people to abandon repetitive and rote work (Ernst, Merola & Samaan, 2019). Thus, opportunities began to arise for the development of new technologies. With the discovery of writing, humanity began to keep records and follow objects, tools, agricultural areas, and even people with these records. When the age of digitalization began, there was a need to transfer this written or processed information to the computer, which led to the emergence of the concept of Optical Character Recognition (OCR). Optical character recognition (OCR) is an excellent engineering tool based on computer vision technology. The software or logic behind OCR is to detect the alphanumerical characters and symbols from the sources of text files, documents, images, etc. This technique is encountered in all areas of our lives and makes human life easier. For example, sometimes we come across the concept of OCR when creating a digital copy of a book edition in a library, and sometimes when recording the address on the letter in a post office. Most importantly, this technology opens the gates of digitalization and automation in the industry.

Especially in the manufacturing sector, part and id numbers are crucial for logistic, regulation, and warranty operations (Chiacchio et al., 2022). Besides, there are different common methods used to process the serial and id codes on the products, namely, screen printed, engraved, and embossed. For instance, when consumer electronics is considered, text characters are screen printed on metal or plastic plates, while in the automotive sector, engraved characters are observed for vehicle id. A vehicle identification number (VIN) is generally engraved on vehicle chassis or metal plates. Embossed characters are usually seen on rubber and plastic parts.

Lots of unexpected setbacks can occur when OCR is performed on production materials and parts, because environmental effects in production, i.e., paint, rust, contamination, etc. may affect the performance of OCR. Besides, when the studies focus on reading engraved or embossed characters reading, these environmental effects are real challenges because OCR methods perform with high accuracy on well-printed texts with known fonts in particular sizes, especially on the plain background which does not lead to low contrast problem (Breithaupt, 2001). When the automotive and tire industries are considered, engraved numbers and letters are used in a variety of cases. In the case of engraved numbers, there is no difference between character color and background color, which leads to low contrast problems. There occurs a mix-up of reflected light and scattered light. Color is not consistent across character areas or background areas (Patil & Dhanvijay, 2015). These aspects of engraved numbers, letters, and characters make conventional OCR approaches not suitable for engraved numbers. These difficulties are rising due to various factors, such as the variation of the light intensity, alignment of text, font size, camera angles, and most importantly surface color, noise, and quality (Zhang, Zhao, Song & Guo, 2013).

This study focuses on the performance comparison of the OCR technique and deep learning-based object detection methods on cured rubber compound specimens. As it is mentioned above, there is no color difference between the surface and characters, which is why traditional OCR cannot be performed at desirable accuracy levels on rubber compound surfaces. Before analyzing the study, the success comes from understanding the main idea of optical character recognition, object detection, and their challenges. Afterward, state-of-the-art object detection methods' performances are compared with OCR on the problem will be stated.

1.1. Literature review

As the name suggests, OCR is a recognition study. This recognition can be achieved through image processing, segmentation, feature extraction, and classification. Basically, optical character recognition is a classification method that classifies and recognizes optical patterns in a digital source corresponding to alphanumeric or other characters (Nagy, Nartker & Rice, 1999; Arica & Yarman-Vural, 2001). This automation approach, OCR, has increased attention in both academic and industrial areas. Although many commercial systems for performing OCR exist for a wide variety of applications, the available solutions are still not able to compete with human reading capabilities with desired accuracy levels, because the human eyes and brain are incredibly adaptive to the changes in surroundings.

The idea of OCR existed to replicate humans' best function, reading, and detecting, by machines. The origins of character recognition come from 1870 when C.R. Carey of Boston Massachusetts invented a retina scanner which was an image transmission system using a mosaic of photocells. Early versions needed to be trained with images of each character and worked on one font at a time (Schantz, 1982; Arica & Yarman-Vural, 2001; Chaudhuri, Mandaviya, Badelia, & Ghosh, 2017). Then, in the early 1900s, a Russian scientist tried to develop an aid for the visually handicapped to read (Mantas, 1986). The first character recognizers appeared in the middle of the 1940s with digital computing developments (Chaudhuri, Mandaviya, Badelia, & Ghosh, 2017). The pioneering work on the automatic recognition of characters has been concentrated either upon the machine-printed text or a small set of well-distinguished handwritten text or symbols, not covering several languages and fonts. Up to this point, machine-printed OCR systems mostly used template matching in which an image is compared to a library of images. Besides, early examples had success in only Latin characters and numerals. However, by emerging digital computing and developments in some sort of complicated algorithms, complicated alphabets were also included in OCR studies such as Japanese, Chinese, Hebrew, Indian, Cyrillic, Greek, and Arabic characters, and numerals (Chaudhuri, Mandaviya, Badelia, & Ghosh, 2017). When It came to the mid-1950s, a new era began. The first OCR solution was commercialized in Reader's Digest, an American general-interest family magazine, in 1954. This technology, OCR, was used to convert typewritten sales reports into punched cards for input into the computer, which replaces human activities (Schurmann et al., 1992).

In the years that followed, large industrial companies began to quickly join the OCR scene. Big players such as IBM, Toshiba, and Hitachi pioneered these studies. The first standardization was done by Americans to increase success in performance, and it is called OCR-A Standards. The big jump in OCR systems was achieved during the 1990s using the new development tools and methodologies which are empowered by the continuously growing information technologies. Mankind has learned that image processing and pattern recognition techniques were efficiently combined with artificial intelligence methodologies such as feed-forward neural networks and recurrent neural networks, which supported high performance on OCR examples (Chaudhuri, Mandaviya, Badelia, & Ghosh, 2017).

Tesseract is one of the well-performed open-source OCR engines which developed at HP between 1984 and 1994 (Rice, Jenkins & Nartker, 1995). The tool contains its pipeline which includes adaptive thresholding, connected component analysis, line and word-finding, and recognition with 2 passes (Smith, 2007). Up to now, the Tesseract engine continues to improve with state-of-the-art techniques and adopt deep learning models in its architecture. In version 4 (v4), a recurrent neural network which is the long short-term memory (LSTM) has been added to the Tesseract OCR engine architecture which yields more accurate results than the non-deep learning models in previous versions of Tesseract (Tesseract user manual, n.d.). Tesseract also permits an opportunity to train with a particular data set to reconfigure the weights of the LSTM which enables the update of the engine to new fonts according to the corresponding study case. This study covers Tesseract v4 OCR engine results.

1.1.1. Challenges in OCR

The performance of OCR applications is affected by both environmental and hardware factors. At this point, various challenges must be dealt with to increase the performance and sustain the system. These difficulties and challenges are briefly described below (Hamad & Mehmet, 2016).

1.1.2. Scene Complexity

Every object outside of the text to be detected poses a challenge to OCR algorithms. The presence of scene complexity in the photographs or images used may result in reduced OCR performance. This complexity brings additional difficulties to users (Ye & Doermann, 2014).

1.1.3. Lighting

Capturing images in real and unconditioned environments often causes uneven lighting and shadows. This poses a challenge to OCR as it distorts the desired properties and features of the image and therefore results in less accurate detection, segmentation, and recognition (Ye & Doermann, 2014; Hamad & Mehmet, 2016).

There are different solutions to prevent uneven lighting. The first of these is to isolate the photographic environment and minimize the environmental effects. Another one is to change the image acquisition method. Taking scanned images of a particular region rather than taking an area image would be an application that increases the performance. Different camera images should be preferred because of the absence of such inequalities in lighting and shadows, better features, and quality of the images taken. For example, using a flashlight or external lighting in addition to the camera we will use while taking the image of a region as a field creates new challenges, although it eliminates such problems with uneven lighting.

1.1.4. Skewness

For optical character recognition systems, the position and condition of the text in the images taken with the camera are very important. If there is a distortion, skewness, or rotation in the image caused by taking the photo, OCR performance will be negatively affected. Additional referencing algorithms are needed for the success and stable operation of the algorithm.

1.1.5. Blurring

The change in the working distance of the camera focus system is an important factor. Character sharpness is the most important parameter in the success of character recognition and character segmentation. At large apertures and short distances, uneven focus can be observed when a small angle of view is changed. There are two types of ambiguity commonly associated with photography: out-of-focus darkness and motion ambiguity (Hamad & Mehmet, 2016). Especially the movement of the object to be photographed or the environmental vibrations in the camera system create blur in the photographs. Therefore, when choosing a camera, attention should be paid to whether the object is moving or not and environmental vibration conditions.

1.1.6. Aspect Ratios

The performance of the OCR algorithm used to detect a text depends on the location, and scale of the text. To simplify these processes, attention should be paid to the aspect ratio of the text on the images taken and it is recommended that it be in a limited variability.

1.1.7. Tilting

The images obtained by conditioned cameras or scanners are generally parallel to the sensor plane, but there may be a tilted view relative to the sensor plane in the images taken on modular cameras or moving systems. Accordingly, lines of text farther from the camera appear smaller than those closer to the camera. This causes the pictures to be skewed. Observing a perspective distortion is obvious if the recognizing perspective is not intolerant, resulting in a lower recognition rate and accuracy (Moravec, 2002).

1.1.8. Fonts

Italic style and script fonts of characters might overlap each other, making it difficult to perform OCR processes such as segmentation (Hamad & Mehmet, 2016). Besides, it is too difficult to perform accurate recognition when the character class number is large with special fonts.

1.1.9. Multilingual Environments

Languages such as Japanese, Chinese, Korean, Hindi, and Arabic have multiple character classes. These languages are quite different from the common Latin characters. Since OCR is more laborious in multilingual situations, and complex symbolism, OCR studies on such sources are still a subject of research today (Smith, Antonova & Lee, 2009).

1.1.10. Deep Learning-Based Object Detection

Object detection is one of the most popular topics in computer vision. Many researchers put hard effort into the subject to develop faster, easy to train, and more accurate object detection methods. After the great success of AlexNet (Krizhevsky, Sutskever & Hinton, 2012) in the 2012 ImageNet classification competition (Deng et al., 2009), convolutional neural networks have been used widely in classification, localization, segmentation, and object detection studies. The great difference between object detection and classification methods is that object detection also predicts the coordinates of a rectangular box that contains the detected object in it. When the location of the detected object is important in a task, object detection algorithms are used. There are roughly two types of object detection methods, namely, one-stage object detectors and two-stage object detectors. Essentially, they differ in accuracy and inference time. The most popular two-stage detectors are R-CNN (Girshick, Donahue, Darrell & Malik, 2014), Fast R-CNN (Girshick, 2015) , Faster R-CNN (Ren, He, Girshick & Sun, 2015), and Mask R-CNN (He et al., 2017), whereas most popular one stage detectors are RetinaNet (Ren, He, Girshick & Sun, 2015), YOLO (Redmon, Divvala, Girshick & Farhadi, 2016), YOLOv2 (Redmon & Farhadi, 2017), YOLOv3 (Redmon & Farhadi, 2018), YOLOv4 (Bochkovskiy, Wang & Liao, 2020), EfficientDet (Tan, Pang & Le, 2020), and YOLOv5 (Jocher et al.,2021).

1.1.11. Challenges of Object Detection

As with any technical problem, there are many difficulties with object detection. These difficulties can principally be listed as real-time prediction speed, data insufficiency and imbalance, and small object detection complication. While addressing the problem desired to be solved with various remedies to these difficulties, it is necessary to establish the system in a way that does not cause weakness by realizing its existence from the very beginning.

1.1.12. Real-Time Detection Speed

For real-time detection, detection speed is also very important along with detection performance. For the cases such as autonomous driving, detection in a flowing conveyor or process, people and asset counting or tracking, etc. detection speed is vital (Srisuk et al., 2019). To overcome this difficulty, fast methods have been developed that can run on simpler hardware.

1.2. Related studies

The performance of the neural network-based method strongly depends on the trained dataset. Models are prone to overfit when they are trained with insufficient datasets (Oksuz, Cam, Kalkan & Akbas, 2020). The imbalance in the amount of data between classes also causes each class not to be predicted with the same performance. The more data is included in the training, the higher performance of the neural network-based object detection algorithm has.

Trying to detect objects of different sizes within the analyzed scene has been a major problem for object detection (Li, Lin, Bai, Li & Yu, 2019). Different anchor structures and feature pyramid networks (Lin et al., 2017) are used to overcome the low detection performance of small objects.

1.3. Conceptual framework

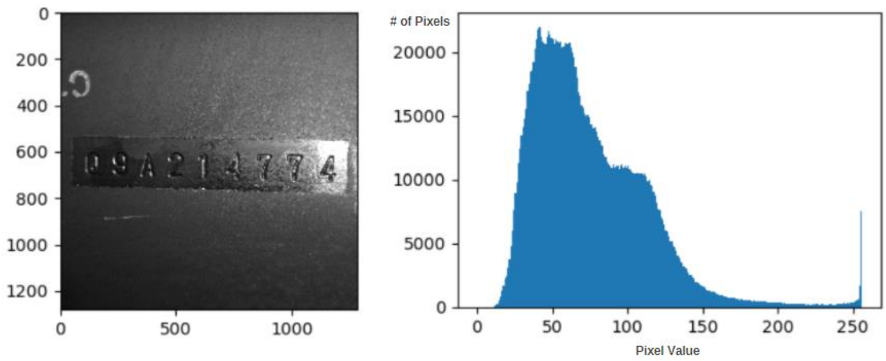
Rubber-based mixtures are used in many industrial products and intermediate products such as hoses, tires, conveyor belts, insulation materials, seals, vehicle parts, shoe soles, and gloves. Carbon black is added to natural rubber or synthetic rubber to enhance the physical properties of the rubber compound. This causes rubber-based mixtures to turn completely black.

As in many manufacturing sectors, the production of rubber-based materials is recorded by various methods to be easily followed during or after the production line. Hence, there is a need to apply serial numbers to finished or semi-finished products for traceability. There are different common methods to process these codes which include alpha-numerical characters and symbols on the products such as screen printed, engraved, and embossed.

The corresponding study covers the detection of applied serial numbers to the rubber compound specimens or products. As mentioned before, rubber compounds have a black surface in their nature, and the detection of these serial numbers turns into the main problem which is to detect black objects on black background. Overall appearance and the serial numbers desired to be detected can be seen in Figure 1. When the histogram of the image has been examined, it is easily seen, as expected, that a right-skewed and narrow distribution of pixel values indicate the low contrast problem which is a well-known technical problem for computer vision.

Figure 1

The serial number on the finished product (left) and histogram for the pixel values (right)



Consistently, it may not be possible to establish a stable image acquisition system in many manufacturing lines as in the corresponding study. It is aimed to make detection via operators' hand terminals or work tablets by capturing the image of the concerned area of the rubber compound or product itself. In this case, low contrast is not the only problem but also perspective, skewness and illumination are other problems. Furthermore, there is natural noise on the surface due to different ingredients in the rubber compound such as oils, accelerators, fillers, etc. which makes detection challenging and reveals the need for a pipeline consisting of specific pre-image processing steps for the high accuracy detection algorithm whether an OCR engine or object detection algorithm.

Figure 2

Several examples of serial numbers on the rubber compound

Materials and Methods

2.1. Data collection

Firstly, to observe the case, a data set is needed to be collected. The data set has been prepared from products and specimens in the production line where the controlling system will be set. Since the desired system is based on capturing the location of the serial numbers via hand terminals or tablets, all the images are taken as replication of the actual instance which may occur when the system will be set. The dataset contains various 5000 images. The objects in the dataset are formed of 14 different types (10 numbers, 4 letters). These are 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, Y, R, A, and Q. All the classes in the images were manually classified and their bounding boxes were labeled for the object detection algorithms. For the fine-tuning of the Tesseract engine, the same dataset has been used.

2.2. Analysis

The study covers two different approaches which are OCR and object detection. To decide which method satisfies the requirements, detailed analyzes were made along with the detection performances and their robustness against scene changes that may occur due to the nature of the problem. When using the OCR method, a problem-specific image processing pipeline is used, while object detection methods are trained based on specific datasets and are used end-to-end without the need for any preprocessing.

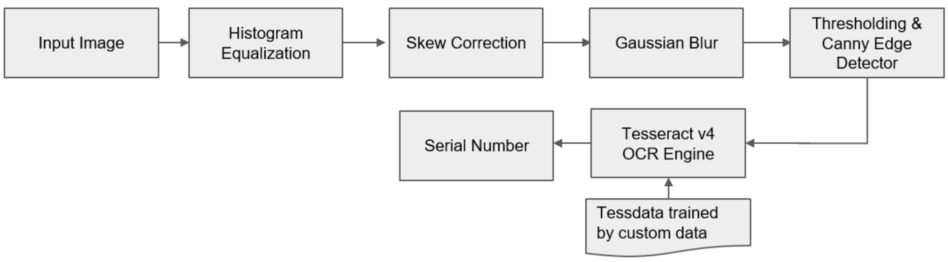
2.3. Procedure

2.3.1. Optical Character Recognition (OCR)

OCR methods are widely used to read documents, scanned files, and texts on products. However, since the nature of the OCR, it is limited to detecting the unfamiliar fonts and characters on its own. Because of this fact, problem-specific preprocess steps are necessary to apply. For the corresponding problem, a pipeline shown in figure 3 is developed. In the proposed pipeline, we try to overcome the challenges of the OCR. Firstly, an arbitrary image is taken as input then it passes through an adaptive histogram equalization (Pizer et al., 1987)to enhance the contrast of the input. After the skew correction is done, gaussian blur is applied to reduce the noise. Then, the Canny Edge Detection algorithm (Canny, 1986) is affixed to improve the accuracy of the Tesseract OCR engine. Tesseract v4 engine is trained by the custom specific data which is already passed through before the pipeline and the tessdata file which includes new weights of its LSTM network is gathered to input the Tesseract v4 engine in the pipeline. Subsequently, a string output has been taken from the pipeline.

Figure 3

Proposed OCR detection Pipeline

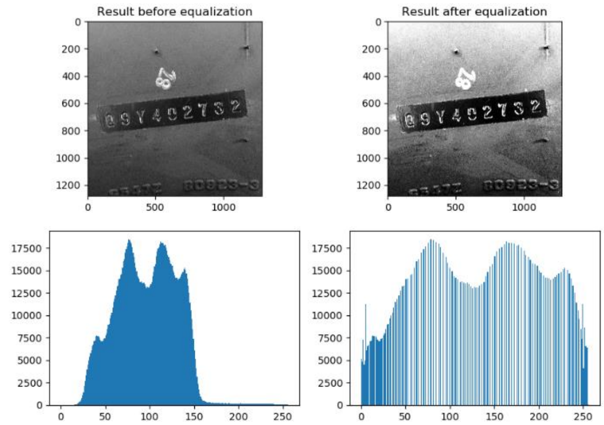


2.3.2. Adaptive Histogram Equalization

Adaptive histogram equalization is an outstanding image processing method to enhance the contrast of an image. To distribute the pixel values in a wide range rather than squeeze between a specific narrow range, adaptive histogram equalization is applied to the taken image as Figure 4.

Figure 4

Original Image (left) and Image after Adaptive Histogram Equalization (right)

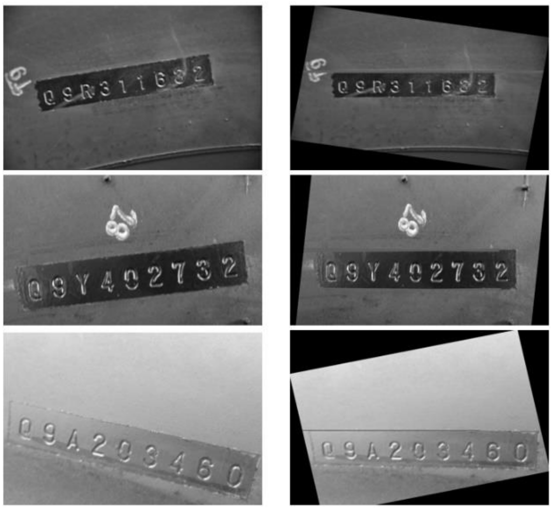


2.3.3. Skew Correction

Skewness is one of the main stability problems in OCR systems. Even though the Tesseract engine can handle low angle skews, in our case samples are coming up to 45 degrees skew. Hence, a skew correction algorithm is needed to be utilized. The algorithm used consists of detecting the straight lines of the serial number area by using the Hough line detection method (Hough, 1962) eliminating the vertical lines and taking the average of the straight lines' angles. Afterward, the image is rotated inverse direction with the magnitude of the average angle value gathered before.

Figure 5

Original Images (left) and Images After Skew Correction (right)

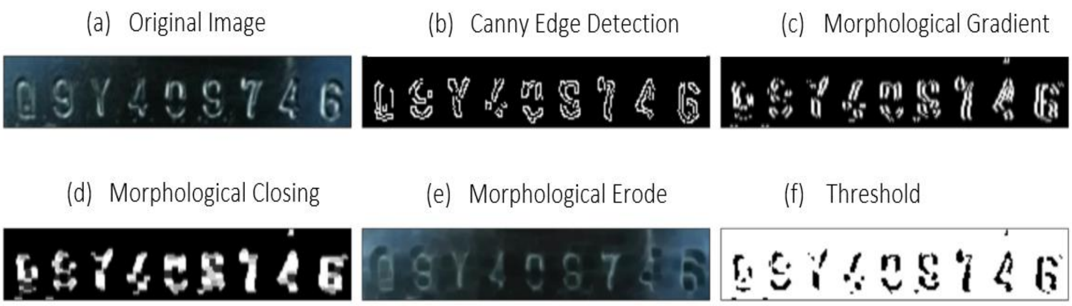


2.3.4. Morphological Operations and Edge Detection

Morphological operations are used to uphold the fundamental context of an image, simplify the data in it and denoise the image. Mathematical morphology enables the use of a toolkit for extracting essential components that may be convenient in the depiction and characterization of region shapes in the image (Serra & Soille, 2012). Dilation, erosion, opening, and closing are commonly used techniques used in image processing. Moreover, Canny Edge Detection is also a good method to eliminate the noise and reveal the edge boundaries of the given object. In figure 6, illustrations of morphological operations and canny edge detection have been given. In the pipeline, canny edge detection and thresholding have been used since morphological operations do not help extract the main features for OCR in the corresponding study.

Figure 6

Morphological Operations (a) Original Image, (b) Canny Edge Detection, (c) Morphological Gradient, (d) Morphological Closing, (e) Morphological Erode, (f) Threshold



2.3.5. Tesseract v4 Engine

Tesseract is one of the most used open-source OCR engines that allows developers to train with custom data. Also, by using the LSTM-based OCR engine, Tesseract v4 provides a wide range of flexibility. In the corresponding problem, a custom training data set is prepared to train the Tesseract to improve the efficiency of OCR performance. The procedure is to set the base model as tessdata best and set the TESSDATA_PREFIX. After setting the training data, LSTM training takes place in Linux operating system. This procedure's output is a tessdata file which is the new weights of the neural network of the Tesseract. It is used in the Tesseract OCR engine in our problem. Moreover, Tesseract provides page segmentation modes (PSM) which assist to improve the quality of the output. In this case, PSM is set to option 6 which assumes the text as a single uniform block.

Despite the entire image processing pipeline and fine-tuning by training on custom data, the OCR solution cannot provide high performance and robust results for the current problem. Both its low accuracy and high inference time have led us to look for an alternative method other than the use of OCR.

2.3.6. Object Detection

Deep learning-based object detection methods output a rectangular bounding box containing the object and its class which is detected by using the convolutional neural networks in the backbone as a feature extractor. To find a solution, a problem-specific custom dataset was created, and neural network training was conducted with this dataset. Enough data for artificial neural network training directly affects the performance and accuracy of the system. Therefore, augmentation studies to increase the collected data set were used before training the object detection algorithms. Afterward, a post-process to record the serial number is applied to the object detection output.

2.3.7. Data Augmentation

The data augmentation approach is adopted to increase the number of samples. Objects obtained from the original images have been synthetically located in the new frames with some variations in their rotations, scales, and translations. For augmentation two kinds of techniques have been used. First, the brightness, contrast, hue, saturation, and noise of images are adjusted differently. Second, random scaling and rotating are applied. After the data augmentation, the total number of objects is increased to 59861 while preserving the distributions of the classes. The distribution of the data set is given in Table 1 below.

Figure 7

Data Augmentation Samples

Table 1

Training Data Distribution

Class	Number of Original Data	Number of Augmented Data	Number of Total Data
Q	1735	4338	6073
R	359	898	1257
Y	356	890	1246
A	1195	2988	4183
0	1956	4890	6846
1	1441	3603	5044
2	875	2188	3063
3	1862	4655	6517
4	1247	3118	4365
5	924	2310	3234
6	1013	2533	3546
7	944	2360	3304
8	796	1990	2786
9	2399	5998	8397

2.3.8. Training

Three basic object detection methods were used with the generated data set: Faster R-CNN (Ren et al., 2015), YOLOv3 (Redmon & Farhadi, 2018), and YOLOv5 (Jocher et al., 2021). All methods were used by optimizing their hyperparameters over their pre-trained weights. The data augmentation features within the models are disabled since data augmentation has been applied to the dataset. Model training with prepared custom datasets was made on a system with NVIDIA GeForce RTX 2080Ti GPU, Intel i7-8700 CPU, and 32 GB RAM with a batch size of 16 for each model. Training lasts approximately 500 epochs for each method.

Figure 8

Output Examples of Object Detection

Results and Discussion

This study aims to detect the engraved letter and number combinations from the surfaces of cured rubber specimens and rubber products. For this purpose, two main approaches which are OCR and object detection are examined. The proposed methods were tested on the same validation set consisting of 500 images. While evaluating the performances of these two methods, detection's mean average precision (mAP) and inference time were considered.

Despite the image process pipeline, the OCR performance was lower than end-to-end object detection methods, since the images on which the detection was desired came at different angles, in different lights, and under variable conditions. In other words, the object detection methods outperform the OCR method with the proposed image process pipeline specific to the problem. Among the object detection methods, YOLOv5 is both more accurate and faster than YOLOv3 and Faster-R-CNN. The results of this study are illustrated in table 2 in aspects of the performance metrics, namely, mean average precision (mAP), recall, F1 score, and inference time.

Table 2

Results of the Methods

Method	mAP	Recall	F1 Score	Inference Time (sec)
Faster R-CNN	0,93	0,84	0,88	0,52
YOLOv3	0,94	0,85	0,89	0,21
YOLOv5	0,95	0,87	0,90	0,14
Tesseract v4 OCR	0,19	0,18	0,19	4,40

4. Conclusion

The motivation behind applying this serial coding is that the traceability of the product is critical for the manufacturing industry. In a conclusion, object detection has been the method to overcome the inherently low contrast problem and the variability of environmental conditions problem among the tried and tested methods for detecting serial numbers on the surfaces of cured rubber specimens and rubber products. YOLOv5 has 0,95 mAP and 0,14 seconds inference time which is quite satisfying for the study. For future study, the output from the object detection can be passed through a recurrent network to predict the serial number in a sequence-to-sequence model manner.

Unfortunately, the performance of OCR methods has not been found as satisfactory. When performances of Faster-RCNN and YOLO algorithms trained with specific-own datasets have been compared to OCR techniques, object detection methods have outperformed OCR methods. For recommendation, the output from the object detection can be passed through a recurrent network to predict the serial number in a sequence-to-sequence model manner for future work.

Acknowledgments

The authors would like to thank Brisa Bridgestone Sabancı Tire Manufacturing and Trading Inc., Turkey for supporting this study. It is a big opportunity that the science behind OCR and object detection to meet real-world production problems. Additionally, the authors would like to declare that some explanations are masked due to industrial confidentiality.

References

Arica, N., & Yarman-Vural, F. T. (2001). An overview of character recognition focused on off-line handwriting. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, *31*(2), 216-233.

Bochkovskiy, A., Wang, C. Y., & Liao, H. Y. M. (2020). Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*. https://arxiv.org/abs/2004.10934

Breithaupt, M. (2001). Improving OCR and ICR accuracy through expert voting. *Computing System Innovations*, 1-16. http://www.csisoft.com/assets/wp_improving-ocr-and-icr.pdf

Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6), 679-698. https://ieeexplore.ieee.org/abstract/document/4767851/

Chaudhuri, A., Mandaviya, K., Badelia, P., & Ghosh, S. K. (2017). Optical character recognition systems. *Optical Character Recognition Systems for Different Languages with Soft Computing. Springer, Cham, 2017. p. 9-41.* https://link.springer.com/chapter/10.1007/978-3-319-50252-6_2

Chiacchio, F., D'Urso, D., Oliveri, L. M., Spitaleri, A., Spampinato, C., & Giordano, D. (2022). A Non-Fungible Token Solution for the Track and Trace of Pharmaceutical Supply Chain. *Applied Sciences*, *12*(8), 4019. https://www.mdpi.com/2076-3417/12/8/4019

Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition* (pp. 248-255). Ieee. https://ieeexplore.ieee.org/abstract/document/5206848/

Ernst, E., Merola, R., & Samaan, D. (2019). Economics of artificial intelligence: Implications for the future of work. *IZA Journal of Labor Policy*, *9*(1). https://sciendo.com/es/article/10.2478/izajolp-2019-0004

Girshick, R. (2015). Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 1440-1448). http://openaccess.thecvf.com/content_iccv_2015/html/Girshick_Fast_R-CNN_ICCV_2015_paper.html

Girshick, R., Donahue, J., Darrell, T., & Malik, J. (2014). Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 580-587). http://openaccess.thecvf.com/content_cvpr_2014/html/Girshick_Rich_Feature_Hierarchies_2014_CVPR_paper.html

Hamad, K. A., & Mehmet, K. A. Y. A. (2016). A detailed analysis of optical character recognition technology. *International Journal of Applied Mathematics Electronics and Computers*, (Special Issue-1), 244-249. https://dergipark.org.tr/en/pub/ijamec/issue/25619/270374

He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision* (pp. 2961-2969). http://openaccess.thecvf.com/content_iccv_2017/html/He_Mask_R-CNN_ICCV_2017_paper.html

Hough, P.V.C. (1962). Method and means for recognising. *US-Patent 3 069 654,1960.*

Jocher, G., Stoken, A., Borovec, J., Christopher, S. T. A. N., & Laughing, L. C. (2021). Ultralytics/yolov5: V4. 0-Nn. silu () Activations Weights & Biases Logging Pytorch Hub Integration. *Zenodo*.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*, 1097-1105. https://proceedings.neurips.cc/paper/2012/hash/c399862d3b9d6b76c8436e924a68c45b-Abstract.html

Li, H., Lin, K., Bai, J., Li, A., & Yu, J. (2019). Small object detection algorithm based on feature pyramid-enhanced fusion SSD. *Complexity*, *2019*. https://www.hindawi.com/journals/complexity/2019/7297960/

Lin, T. Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2017). Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2117-2125). http://openaccess.thecvf.com/content_iccv_2017/html/Lin_Focal_Loss_for_ICCV_2017_paper.html

Mantas, J. (1986). An overview of character recognition methodologies. *Pattern recognition*, *19*(6), 425-430.

Moravec, K. (2002, September). A Grayscale Reader for Camera Images of Xerox DataGlyphs. In *BMVC* (pp. 1-10). http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.64.4555&rep=rep1&type=pdf

Nagy, G., Nartker, T. A., & Rice, S. V. (1999, December). Optical character recognition: An illustrated guide to the frontier. In *Document Recognition and Retrieval VII* (Vol. 3967, pp. 58-69). International Society for Optics and Photonics. https://www.spiedigitallibrary.org/conference-proceedings-of-spie/3967/0000/Optical-character-recognition-an-illustrated-guide-to-the-frontier/10.1117/12.373511.short

Oksuz, K., Cam, B. C., Kalkan, S., & Akbas, E. (2020). Imbalance problems in object detection: A review. *IEEE transactions on pattern analysis and machine intelligence*. https://ieeexplore.ieee.org/abstract/document/9042296/

Patil, A. V., & Dhanvijay, M. M. (2015, December). Engraved character recognition using computer vision to recognize engine and chassis numbers: Computer vision technique to identify engraved numbers. In *2015 International Conference on Information Processing (ICIP)* (pp. 151-154). IEEE. https://ieeexplore.ieee.org/abstract/document/7489368/

Pizer, S. M., Amburn, E. P., Austin, J. D., Cromartie, R., Geselowitz, A., Greer, T., ... & Zuiderveld, K. (1987). Adaptive histogram equalization and its variations. *Computer vision, graphics, and image processing*, *39*(3), 355-368. https://www.sciencedirect.com/science/article/pii/S0734189X8780186X

Redmon, J., & Farhadi, A. (2017). YOLO9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 7263-7271). http://openaccess.thecvf.com/content_cvpr_2017/html/Redmon_YOLO9000_Better_Faster_CVPR_2017_paper.html

Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*. https://arxiv.org/abs/1804.02767

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 779-788). https://www.cv-foundation.org/openaccess/content_cvpr_2016/html/Redmon_You_Only_Look_CVPR_2016_paper.html

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*. https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. *arXiv preprint arXiv:1506.01497*. https://proceedings.neurips.cc/paper/2015/hash/14bfa6bb14875e45bba028a21ed38046-Abstract.html

Rice, S. V., Jenkins, F. R., & Nartker, T. A. (1995). *The fourth annual test of OCR accuracy* (Vol. 3). Technical Report 95. https://raw.githubusercontent.com/tesseract-ocr/docs/master/AT-1995.pdf

Schantz, H. F. (1982). *History of OCR, optical character recognition*. Recognition Technologies Users Association. https://dl.acm.org/doi/abs/10.5555/539391

Schurmann, J., Bartneck, N., Bayer, T., Franke, J., Mandler, E., & Oberlander, M. (1992). Document analysis-from pixels to contents. *Proceedings of the IEEE*, *80*(7), 1101-1119. https://ieeexplore.ieee.org/abstract/document/156473/

Serra, J., & Soille, P. (Eds.). (2012). *Mathematical morphology and its applications to image processing* (Vol. 2). Springer Science & Business Media. https://link.springer.com/book/10.1007/978-94-011-1040-2

Smith, R. (2007, September). An overview of the Tesseract OCR engine. In *Ninth international conference on document analysis and recognition (ICDAR 2007)* (Vol. 2, pp. 629-633). IEEE. https://ieeexplore.ieee.org/abstract/document/4376991/

Smith, R., Antonova, D., & Lee, D. S. (2009, July). Adapting the Tesseract open-source OCR engine for multilingual OCR. In *Proceedings of the International Workshop on Multilingual OCR* (pp. 1-8). https://dl.acm.org/doi/abs/10.1145/1577802.1577804

Srisuk, S., Suwannapong, C., Kitisriworapan, S., Kaewsong, A., & Ongkittikul, S. (2019). Performance evaluation of real-time object detection algorithms. *2019 7th International Electrical Engineering Congress (iEECON), 2019. doi:10.1109/ieecon45304.2019.8938687*

Tan, M., Pang, R., & Le, Q. V. (2020). Efficientdet: Scalable and efficient object detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 10781-10790). http://openaccess.thecvf.com/content_CVPR_2020/html/Tan_EfficientDet_Scalable_and_Efficient_Object_Detection_CVPR_2020_paper.html

Tesseract user manual. (n.d.). *Retrieved May 23, 2021, from https://tesseract-ocr.github.io/tessdoc/*

Ye, Q., & Doermann, D. (2014). Text detection and recognition in imagery: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *37*(7), 1480-1500. https://ieeexplore.ieee.org/abstract/document/6945320/

Zhang, H., Zhao, K., Song, Y. Z., & Guo, J. (2013). Text extraction from natural scene image: A survey. *Neurocomputing*, *122*, 310-323. https://www.sciencedirect.com/science/article/pii/S0925231213006309