Intelligent Methods in Engineering Sciences 2(4): 090-101 2023

IMIENS

INTELLIGENT METHODS
IN ENGINEERING SCIENCES

https://www.imiens.org

International Open Access

December, 2023

e-ISSN 2979-9236

Research Article

https://doi.org/10.58190/imiens.2023.69

Digitization and Archiving of Company Invoices using Deep Learning and Text Recognition-Processing Techniques

Ensar GUNAYDIN [a], ***Bunyamin GENCTURK*** [b], ***Cuneyt ERGEN*** [a], ***Murat KOKLU*** [c, *]

[a] *ALISAN Logistic Inc., Istanbul, Türkiye*
[b] *Technology Faculty, Selcuk University, Konya, Türkiye*
[c] *Department of Computer Engineering, Selcuk University, 42250 Selcuklu, Konya, Türkiye*

ARTICLE INFO

Article history:
Received 10 November 2023
Accepted 27 December 2023

Keywords:
Optical Character Recognition,
Image Preprocessing,
Document Digitization,
Artificial Intelligence,
Deep Learning

ABSTRACT

Nowadays, it is crucial to transfer official documents such as invoices, dispatch notes, and receipts into digital environments and establish correct semantic relationships. However, understanding and processing these documents is a difficult process that requires significant time and effort. In recent years, the use of deep learning, image preprocessing, text detection, and optical character recognition (OCR) technologies have made this process easier. However, for text recognition and processing techniques to produce accurate results, documents must be clean and readable. Additionally, difficulties arising from time-consuming, tiring, error-prone, and cost-incurring human-powered digitalization processes must be reduced. The aim of this study is to digitize and archive scanned invoices and similar official documents using current artificial intelligence technologies, thereby enabling the most effective use of components such as time, cost, and human resources. The dataset used in the study includes 10,000 ".jpg" image files and 10,000 ".xml" data files. The model trained with the ResNet-50 architecture can detect text with accuracy rates of up to 97% on randomly selected images from the dataset. In an environment where a person can process an average of 2,112 documents per month, it is predicted that the trained artificial intelligence model can process 108,000 documents per month. With this developed method, businesses can quickly digitize and archive official documents such as invoices, dispatch notes, and receipts. Future studies propose the development of new methods that can produce better results using larger and more diverse datasets.

1. INTRODUCTION

People create documents to record and protect information. However, reading and understanding documents such as invoices, contracts, and resumes requires time and effort. Therefore, nowadays, it is preferred that documents are stored in electronic environments. However, many documents that require physical processing are still not available in electronic form. After processing, these documents are transferred to electronic environments as scanned images or photos [1, 2]. Detecting text and shapes in scanned documents is considered a difficult problem. With the development of deep learning, the fields of natural language processing and computer vision have made tremendous progress. These developed methods are frequently preferred for understanding and analyzing documents [3-5].

Correctly establishing semantic relationships in official documents such as invoices, delivery notes, and receipts is crucial for focusing on relevant areas. The first step in this process is to be able to detect regions on the image because the text on document images cannot be directly identified and detected by modern search and analytical tools. Text detection technology comes into play at this point and plays an important role in understanding document images and extracting information. Text detection technologies typically take a data-driven approach using a mix of machine-learning techniques and rule-based learning models [1, 6, 7]. Text detection is preferred in finding logical objects with diversity such as text lines, formulas, tables, and shapes in documents. Text regions in complex backgrounds are automatically detected and labeled with bounding boxes. Additionally, automatically processable information is a valuable and sought-after technique for literature mining, domain researchers, and data scientists [8, 9].

Correct labeling of data also increases reading accuracy in many tasks such as Optical Character Recognition (OCR), where recent developments in deep learning have played an important role in tasks such as business card recognition, and license plate recognition [10], or handwriting text recognition. OCR is the process of converting printed or handwritten documents into a

*** Corresponding Author:** mkoklu@selcuk.edu.tr

structured and digitized format. However, OCR may face some challenges in official documents such as invoices and receipts, where higher accuracy is required. Therefore, manual labeling is still widely used [11-14]. To reduce or even eliminate manual work, fast, reliable, and highly accurate OCR is needed to be developed [15].

To achieve high accuracy in text recognition and processing, documents need to be clean and readable. However, scanned documents often have noise, blurriness, ink stains, fading, paper aging, etc., which makes the image quality low and the recognition accuracy poor. Therefore, many important documents are not digitized and were kept offline. Preprocessing of document images is often necessary before OCR [1, 16, 17]. Image preprocessing sometimes plays a crucial role in deciding whether a document image meets OCR requirements or not [18, 19].

The main purpose of the preprocessing stage is to reduce or correct noise and blurriness in document images, and improve the image quality for further processing steps [20]. Image restoration techniques, which are pioneers in this field, have attracted increasing attention in recent years. As the demand for text recognition and processing increases, new images with less noise and less blurriness, and closer to the original noise-free image, are being attempted to be produced [21, 22].

With the development of artificial intelligence technologies, it is expected that companies' information technology departments and Research and Development (R&D) departments will offer innovative solutions. Companies should identify the shortcomings in the field to maintain their competitive advantage and develop solutions and innovative ideas to strengthen their position in the industry. In this context, it is crucial to develop technologies that minimize or ideally eliminate the negative effects in processes such as processing, archiving, and analyzing invoices, delivery notes, and receipts.

2. Related Works

After reviewing the literature, it is observed that methods such as deep learning, image preprocessing, object detection, text recognition and processing are preferred for tracking and control of official documents, especially in recent years. Some of these studies are listed as follows:

Tupaj et al., (1996) proposed an OCR-based table detection technique. In this method, row sequences with table-like characteristics are searched based on key words that can be found in table headers. Rows containing key words are accepted as the starting row, and subsequent rows are categorized and matched as table structure. It is noted that the limitation of this technique is largely dependent on the key words that can appear in table headers.

Zhao et al., (2018) used a method called Skip Connected Deep Convolutional Autoencoder (SCDCA), consisting of multiple convolutional layers, followed by batch normalization layers and leaky rectified linear unit activation functions. Inspired by the idea of skip connections, two types of skip connections are used in the network. The results from the trained model are evaluated using OCR testing to restore the data.

Li et al., (2018) propose a hybrid method that combines deep structured prediction and supervised clustering to detect formulas, tables, and figures in PDF document images in a unified framework. After inference using graphical-based models (CRF), the row regions of the same class within a cluster are grouped into a page-level object. It is noted that if different primitive regions are used instead of line regions for classification and clustering, it can be applied to more complex and irregular documents outside of PDFs.

Saha et al., (2019) presented a new, end-to-end trainable, deep learning-based method called Graphic Object Detection (GOD) to localize graphical objects in document images. The developed method is stated to be data-driven and used to detect graphical objects in document images. The performance analysis carried out on various public datasets such as ICDAR-2013, ICDAR-POD 2017, and UNLV reportedly yielded promising results.

Soni et al., (2019) presented a text detection and localization approach based on a new text-awareness model. They use a fast directional filter to effectively separate connected characters by removing mixed pixels at the edges of blurry images. This filter incorporates an improved fast edge-preserving and smoothing maximum stable extremal region (FEPS-MSER) algorithm. The Naive Bayes classifier used in the study contributes to the classification of text and non-text components using graph cut algorithm by assisting in the accurate and rapid determination of the text-awareness score.

Subramani et al., (2020) stated that text recognition revolves around hand-crafted features to detect characters. It is observed that some general themes in modern NLP and document understanding are discussed, and end-to-end automatic document understanding systems are created. They are concerned with the best methods for both text detection and transcription for OCR. Multiple approaches have been presented for document layout analysis.

Weiwei Sun et al., (2022) proposed an EAST detection algorithm developed for detecting and recognizing inclined text in images. This algorithm uses reinforcement learning to train a recurrent neural network controller. By selecting an optimal fully convolutional neural network structure, the multiscale features of the text are extracted. This information is then passed to the Generalized Intersection over Union (GIoU) algorithm, aiming to increase the regression effect of the text bounding box. The loss function is adjusted to balance the positive and negative example classes before extracting the improved text detection results.

As well as the details of the studies in the literature, summary information is provided in Table 1. Upon examination of Table 1, it can be observed that the number of data used in the studies varies between 144 and 2417. The studies were generally conducted by testing different methods, and the accuracy rates also varied depending on the number of data and the method used.

Table 1. Summary information about the studies found in the literature.

Ranking	Number Of Data	Method	Accuracy (%)	References
1	1470	OCR	71	[23]
2	144	SCDCA	20	[1]
3	2417	CRF	60-80	[8]
4	1500	R-CNN	64	[3]
5	1455	FEPS-MSER	83	[24]
6	1000	R-CNN	-	[2]
7	1962	EAST	84	[25]

As expressed in literature studies, information on official documents such as invoices, delivery notes, and receipts needs to be processed individually and completely digitized. However, the accumulation of past records can cause significant time and personnel losses in institutions due to the need to process data individually and transfer it to digital format.

Information technologies have been studied in recent years, but technology is always changing, so adapting to new technologies is essential. It is observed that official documents such as invoices and delivery notes are generally digitized using human labor, which causes both time and labor losses. By training invoice and delivery note images with deep learning algorithms, high-precision text detection and processing will be achieved with the resulting model, allowing for the most efficient and effective use of human labor and time components.

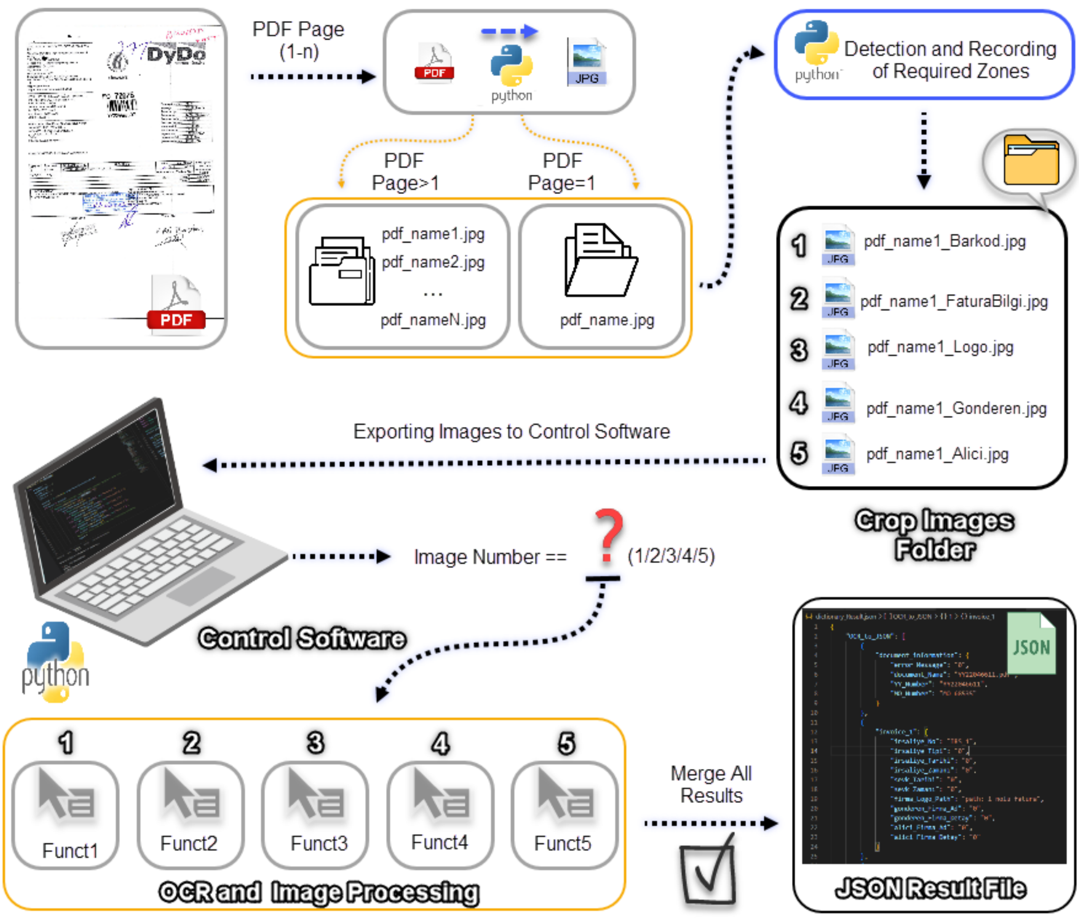


Figure 1. Block diagram illustrating the workflow of the study.

The aim of this study is to present new methods for effectively digitizing and archiving scanned images of invoices and other official documents using current artificial intelligence technologies. Through combining various artificial intelligence techniques, this study contributes to the literature by creating a more efficient method. In comparison to previous studies, this study has the following characteristics:

- Deep learning,
- Object recognition,
- Image preprocessing,
- Optical character recognition

Which these methods operate in a highly interactive and interconnected manner. As a result, productivity increases by making the best use possible of time, cost, and human resources.

3. Materials and Method

This section provides information on the method developed for digitizing and archiving invoice images. Deep learning, image processing, and optical character recognition methods were used in the study, with Python programming language being preferred for implementing these methods. First, a model was trained for region labeling and object recognition on invoice images. Based on the outputs of the trained model, text recognition and processing were performed. The flowchart demonstrating the process is presented in Figure 1. Furthermore, detailed information on the techniques used and the dataset is provided in this section. Following that, the methods used, architectures used, and other relevant details are mentioned in the next sections.

3.1. Image acquisition

In this study, various invoice images belonging to various companies were used for invoice processing. These images were obtained with the permission and approval of the authorized company. After the invoice originals were scanned, they were uploaded to the cloud system in ".pdf" format. These files obtained from the company are converted into ".jpg" format images with the same name as the ".pdf" extension implementing our Python code. This conversion process is shown in Figure 2.

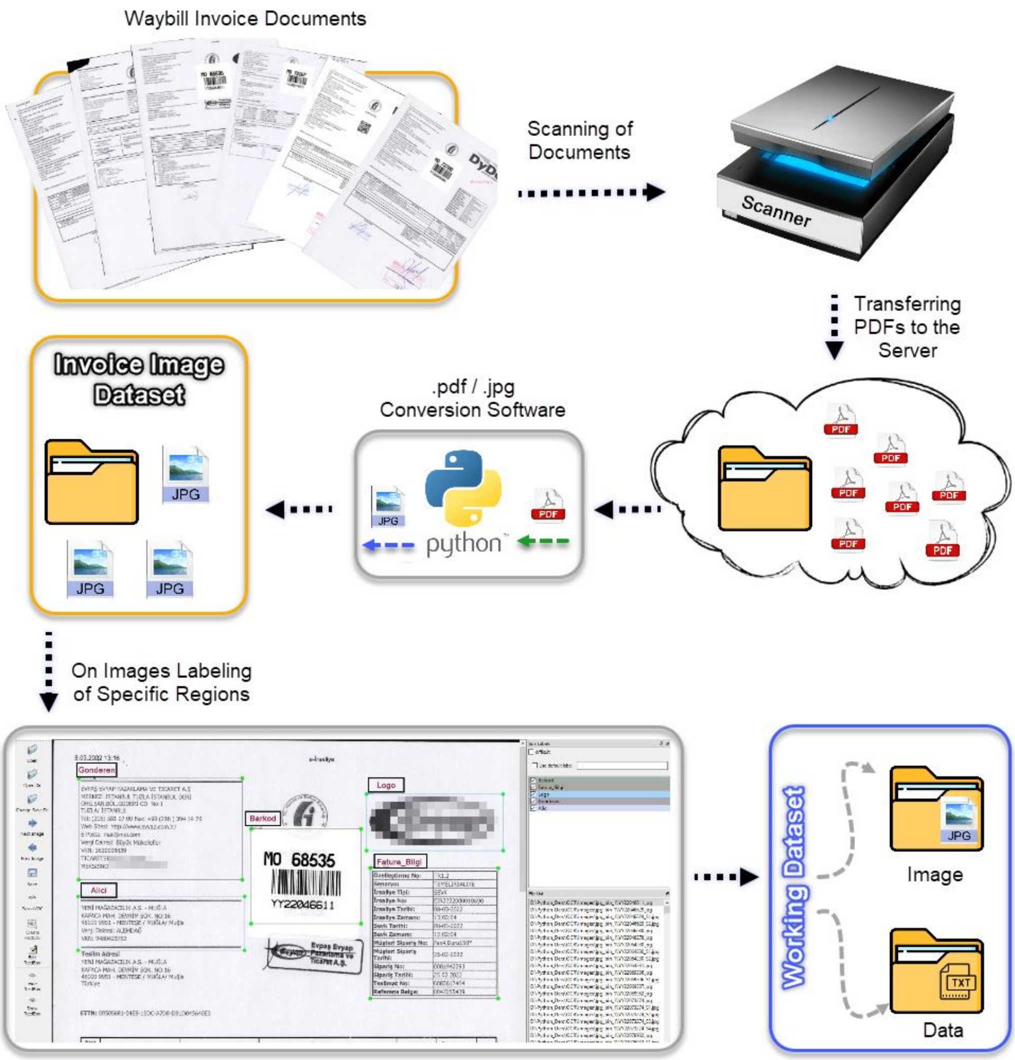
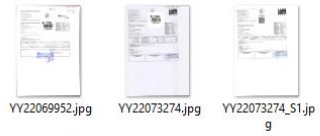
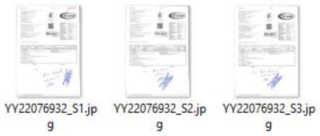
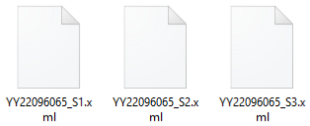
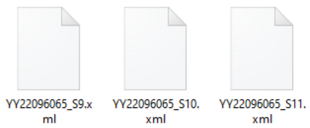


Figure 2. Data acquisition process

3.2. Image acquisition

The dataset consists of a total of 10,000 invoice images and corresponding data files in ".xml" format. As part of our project, we created this dataset. As shown in Table 2, there are 10,000 ".jpg" image files and 10,000 ".xml" data files. The images were used as they were scanned from the original documents and no changes were made to the background.

Table 2. Invoice data information table.

Data Type	Sample Demonstration	Data Count
Image File (".jpg")	 YY22069952.jpg YY22073274.jpg YY22073274_S1.jpg  YY22076932_S1.jpg YY22076932_S2.jpg YY22076932_S3.jpg	10.000
Data File (".xml")	 YY22096065_S1.xml YY22096065_S2.xml YY22096065_S3.xml  YY22096065_S9.xml YY22096065_S10.xml YY22096065_S11.xml	10.000
Total		20.000

The images in this dataset have RGB or gray color space, non-standard pixel sizes, and are in the ".jpg" format. The dataset was randomly divided into 80% training and 20% testing data, which is a commonly used practice in the literature, which is illustrated in Figure 3.

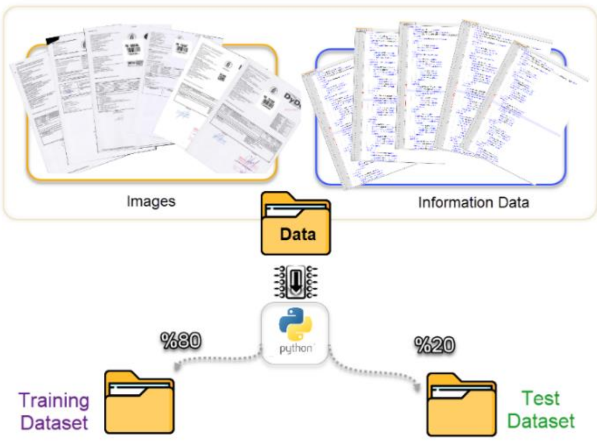


Figure 3. Distribution of data as training and test sets

3.3. Labeling Images and Model Training

Scanned documents typically contain more text than natural images, leading to specific challenges in detecting text in documents. Trying to read unwanted areas with text results in increased workload and time consumption. Various methods are used to minimize these difficulties. An example segmentation method is used to classify each pixel of an image into predefined categories. Pixel-based operations are performed, and these pixel analyses are typically used to estimate the probabilities of text regions and relationships between characters. However, scanned document images contain a lot of noise and unwanted symbols, resulting in less accurate text output than expected. This means that Tesseract, a character recognition engine, cannot output the text 100% accurately. These unwanted situations need to be minimized [2, 26-28]. Labeled images are trained using deep learning methods, as shown in Figure 4, and necessary areas are detected. Thus, only the required areas are sent to OCR, resulting in more accurate results.

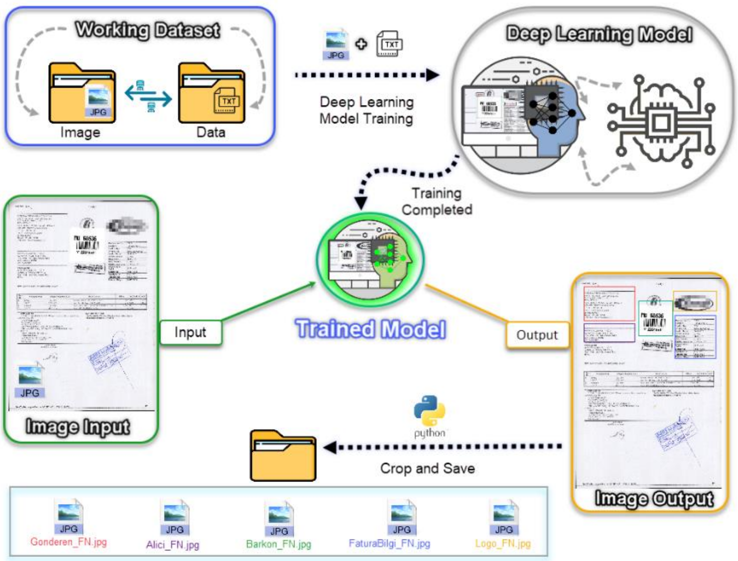


Figure 4. Training the model and block diagram showing the inputs and outputs of the trained model.

3.4. Object Detection

In traditional text detection methods, handcrafted features were used for character detection [29]. However, advancements in fields such as object detection and semantic segmentation in deep learning have brought a new perspective to text detection [30, 31]. Using high-performing object detectors from traditional computer vision literature such as Single Shot MultiBox Detector (SSD) and Faster R-CNN models, efficient text detectors can be created [32-34].

3.5. Convolutional Neural Networks

There are many different architectures of convolutional neural networks (CNN) in the literature. These architectures are developed by taking inspiration from the natural visual perception mechanisms of living organisms [35]. CNN architectures extract features through layers based on input data and learn and classify using these features. Essentially, CNN architectures consist of five layers: convolutional layer, pooling layer, activation layer, fully connected layer, and softmax layer [36-40]. The learning process is one of the most important and challenging stages in CNN architectures. This process is related to finding the most suitable values in the search space for different datasets and these values vary [41-46].

In this study, ResNet-50 was chosen among the convolutional neural network models. ResNet-50 is resistant to overfitting and provides high accuracy compared to similar models. Residual Networks (ResNet), which was first developed by He Kaiming and colleagues in 2015, includes structures with the highest number of layers and has been trained on more than a million images in the ImageNet database [47-49]. ResNet-50 is a convolutional neural network with 50 layers and is commonly used in deep learning [47, 50]. Structurally, ResNet-50 transfers the values required to reach the final result through prediction from one layer to another. To improve the accuracy of these predictions, an Squeeze and Excitation (SE) block structure was preferred, which improves the quality of results generated in a network [47, 51-54].

3.6. Optical Character Recognition (OCR)

OCR consists of two main components: text detection and text transcription. Text detection aims to locate text within a page or image. Input in image format is typically represented as a three-dimensional tensor, C × H × W, where C represents the number of channels (usually red, green, and blue), H represents the height, and W represents the width. As seen in the images provided in Figure 5, text comes in various forms and is often corrupted [2, 55]. Image processing techniques minimize these corruptions and enable OCR to function more successfully.

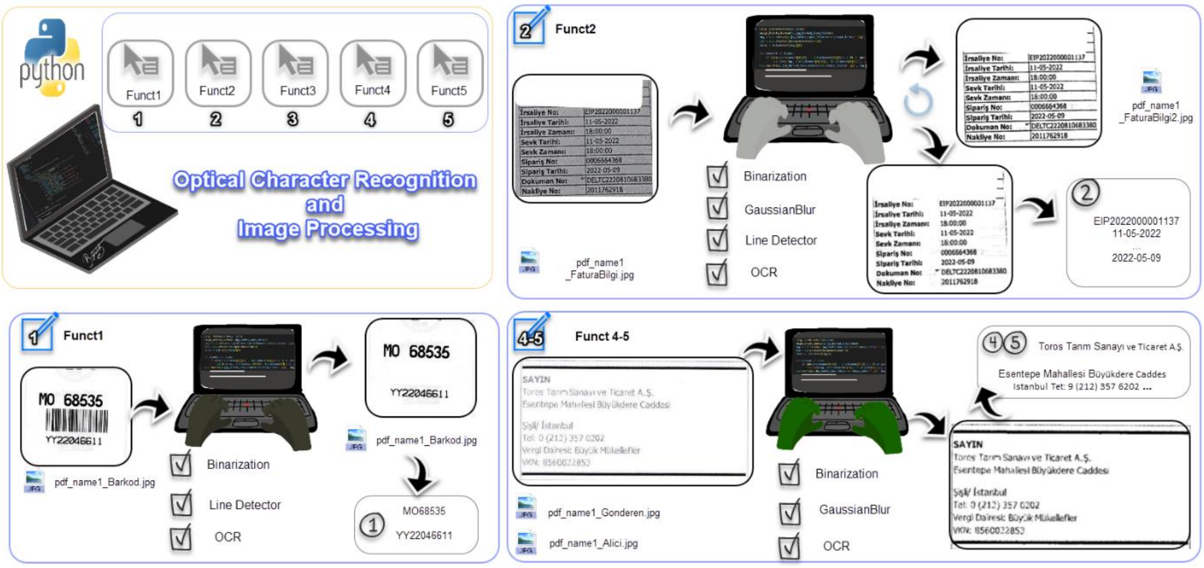


Figure 5. The functions used for OCR in this study are.

In digital image processing, image preprocessing techniques are frequently used. These techniques are used to reduce unwanted noise in the image, correct potential errors, and bring the image to the desired format. Additionally, image preprocessing can be applied to improve the performance of deep learning algorithms. For instance, operations like edge detection, color changes reduction, and background noise removal can be utilized to enhance the accuracy of the algorithm. Therefore, image preprocessing is an important step in digital image processing. It is also necessary to achieve better results in studies [44, 56-58]. This technique has resulted in significant performance and accuracy improvements in the recognition and processing of texts.

4. Experimental Results

In this section, the detection of the necessary regions in official document images, followed by the analysis of the outputs obtained through image processing and text recognition methods, are presented. Additionally, the findings, results, and related details of the study are also included in this section. The model training was conducted in Python programming environment. The system specifications of the computer used in the study are given in Table 3.

Table 3. Computer specifications used for training deep learning models.

Configuration Item	Value
Processor (CPU)	Intel(R) Core (TM) i7-7700HQ CPU @ 2.80GHz
Display Card (GPU)	NVIDIA GeForce GTX 1050
RAM	16 GB
Hard Disk	1TB HDD + 1TB SSD
Operating System	Windows 10 Home Single Language

Additionally, platforms such as Tensorflow, VSCode, Tesseract, LabelImg, etc. were preferred for their positive contributions and conveniences to the study. Operations such as machine learning, image preprocessing, object detection, text recognition and processing were performed with the computer whose specifications were mentioned in Table 3. Optimal classification success rates were calculated for the detection of labeled regions in the images, and it was aimed to use stable results in the application. The parameters used in the training of the deep learning model are given in Table 4.

Table 4. Required parameters used in the training of ResNet-50.

Parameter	Value
Num_classes:	5
Height x Width:	640 x 640
Type:	Ssd_resnet50_v1_fpn_keras
Min_depth:	16
Batch_size:	1
Num_steps:	Min (10.000) – Max (100.000)
Checkpoint_type:	Detection

When the parameters and their values given in Table 4 are examined, Num_classes represent the number of labeled classes. In total, 5 classes are defined, barcode, invoice information, logo, sender, and recipient respectively. The input dimensions of the images to the model were set as 640x640 pixels. This was done to minimize the distortions in the images and save time in training while preserving the text quality. The Batch_size parameter, which represents the number of images kept in RAM during model training, was assigned a value of "1" to achieve optimal performance.

In the deep learning model, the Num_steps parameter, which directly affects performance results, was not fixed in this study. To analyze the success rates in detecting the necessary regions, 3 iteration numbers varying between 10,000 and 100,000 were determined. These iteration numbers are 10,000, 50,000, and 100,000, respectively. Afterwards, the model was exported separately for the 3 iteration values. To evaluate the model's performance at intervals, 2 randomly selected official document images from the dataset were chosen. The results and evaluations obtained from the outputs of the deep learning model are presented in this section respectively.

Object Detection Results of ResNet-50 Model (10k, 50k and 100k iterations):

The selected images from the dataset are named as 'Sample Picture 1 (SP1)' and 'Sample Picture 2 (SP2)' for random selection. Some areas of the images were censored to ensure the confidentiality of company information. As seen in Figure 6, the training of the deep learning model was insufficient and some of the necessary regions could not be detected, leading to various issues. To minimize these problems, the model was trained up to 100,000 iterations and results were observed at certain intervals as illustrated in figures (Figure 6, 7, and 8).

The object detection model trained with 10,000 iterations using the ResNet-50 architecture detects the invoice information and sender labels for SP1 with success rates of 57% and 48%, respectively, as seen in Figure 6. Due to the low iteration count, the barcode, logo, and receiver labels were either undetected or detected with low success rates and are therefore not shown. Similarly, when the same model was examined for SP2, it achieved success rates of 57%, 61%, and 37% for the invoice information, sender, and receiver labels, respectively. Since these rates were considered insufficient for the study, the model was further trained. During the training process, possible issues were monitored by analyzing the classification loss graph for the ResNet-50 model, which is presented in Figure 9.

When examining Figure 9, the classification losses of the model initially decrease linearly. This indicates that the model has successfully learned and is able to detect the necessary regions with better performance. Looking at the rest of the graph, it can be observed that there is no linearity in the classification losses, with fluctuations of increasing and decreasing at various intervals. The classification loss values of the trained model vary between 0.02 and 1.0. It has been determined through the analysis that the lower this value, the more successful the learning process of the model is. Considering the obtained results, it is observed that there are no significant changes in the results after 50,000 iterations. Therefore, the training was completed in this way to avoid the model from overfitting. The success rates resulting from the model being trained at all the specified iteration numbers are individually listed in Table 5.

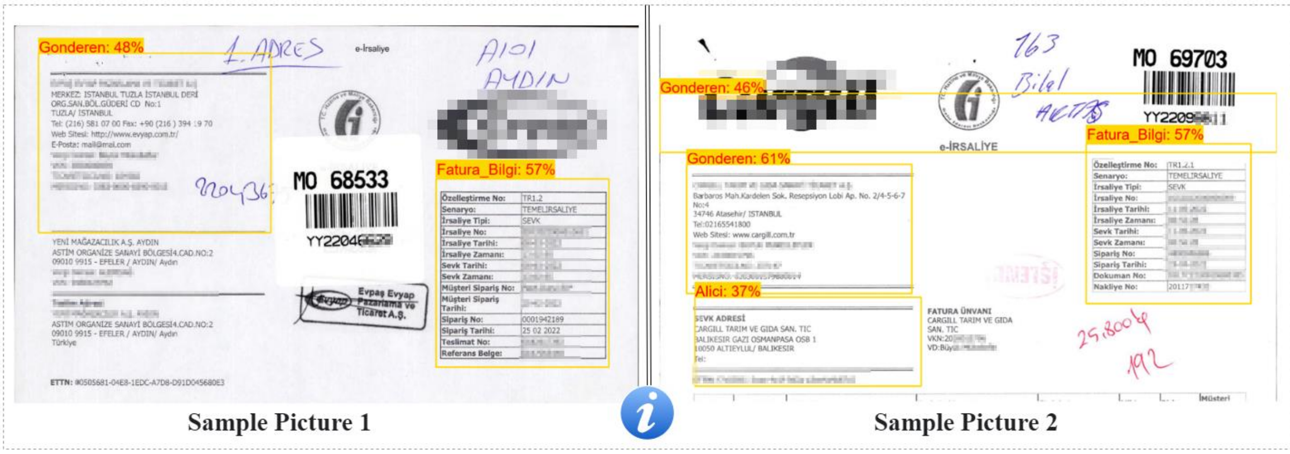


Figure 6. Object Detection Results (10.000 iteration)

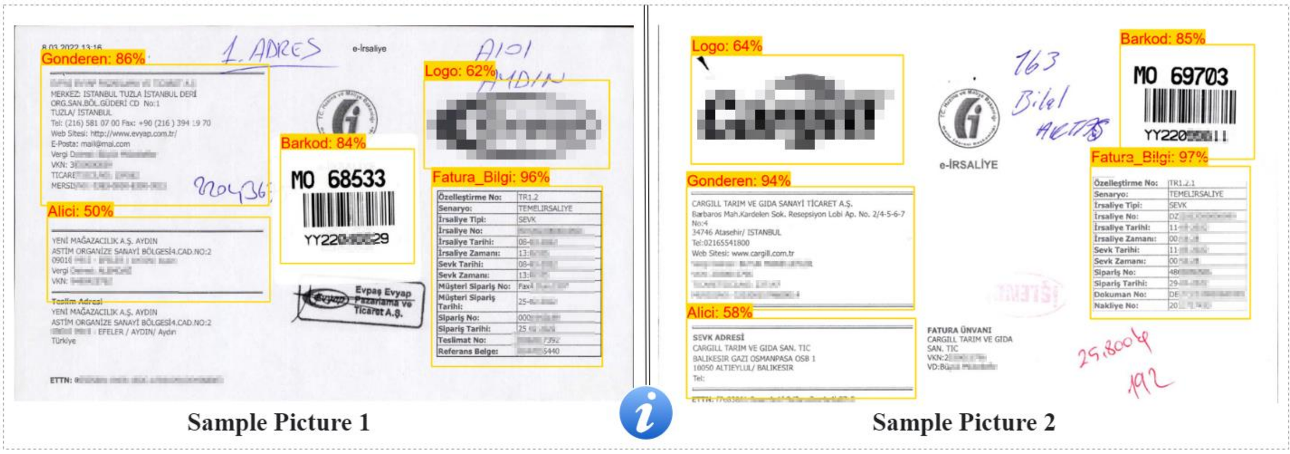


Figure 7. Object Detection Results (50.000 iteration)

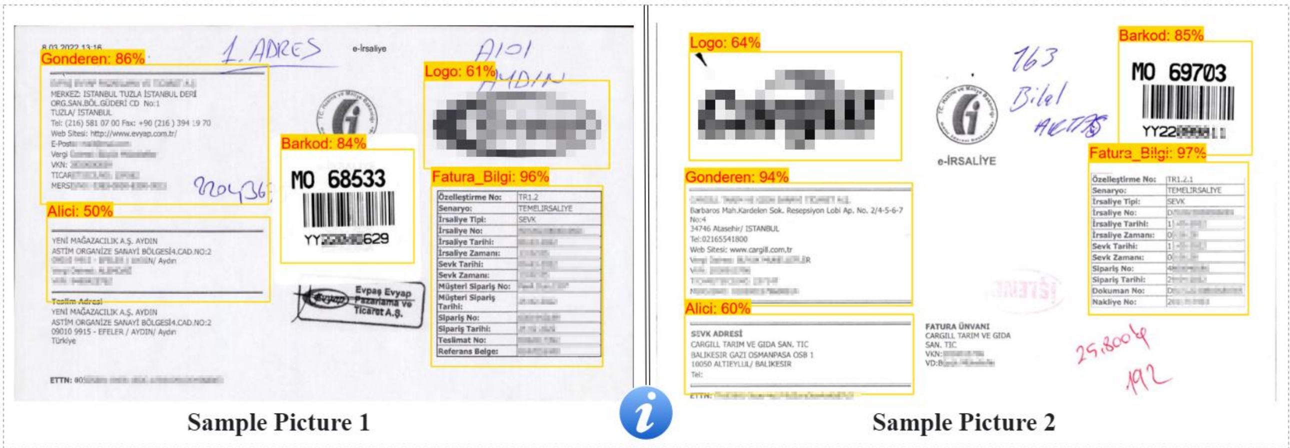


Figure 8. Object Detection Results (100.000 iteration)

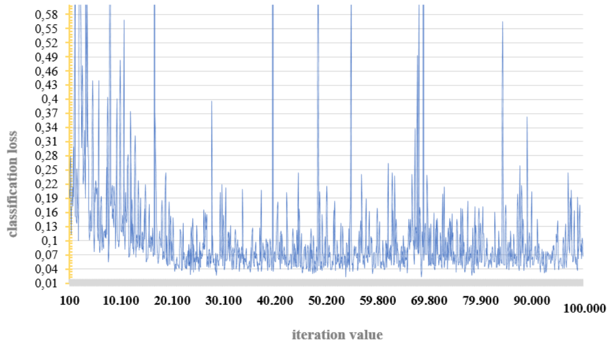


Figure 9. Classification loss graph for ResNet-50 model.

When examining Figure 9, the classification losses of the model initially decrease linearly. This indicates that the model has successfully learned and is able to detect the necessary regions with better performance. Looking at the rest of the graph, it can be observed that there is no linearity in the classification losses, with fluctuations of increasing and decreasing at various intervals. The classification loss values of the trained model vary between 0.02 and 1.0. It has been determined through the analysis that the lower this value, the more successful the learning process of the model is. Considering the obtained results, it is observed that there are no significant changes in the results after 50,000 iterations. Therefore, the training was completed in this way to avoid the model from overfitting. The success rates resulting from the model being trained at all the specified iteration numbers are individually listed in Table 5.

Table 5. The object detection results of the model trained at all specified iteration values.

Class	Sample Picture 1 (%)			Sample Picture 2 (%)		
	10.000	50.000	100.000	10.000	50.000	100.000
Barcode	-	84	84	-	85	85
Invoice information	57	96	96	57	97	97
Logo	-	62	61	-	64	64
Sender company	48	86	86	61,46	94	94
Recipient company	-	50	50	37	58	60

When the iteration number was increased up to 50,000, it can be seen from Table 5 that the accuracies of detecting the regions have also increased. However, it is understood that after 50k, the learning rate of the model generally did not change, and when it did, it showed an increase or decrease in the range of 1-2%. Since these rates are tolerable, they do not directly affect the text recognition and processing results. Of course, it is not enough to detect only these regions for the study. After these regions are detected, they are cropped and saved in the folder directory by the software. Even if these steps are successfully performed, there may still be some problems in the next process. For example, it is inevitable that there will be various corruptions due to scanning or shading problems in the official documents to be uploaded to the system. Therefore, it is necessary to determine and implement techniques that will minimize or eliminate these issues.

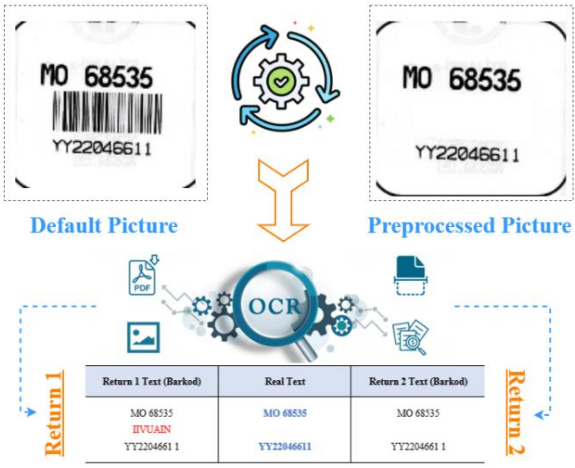


Figure 10. Image preprocessing for barcode label

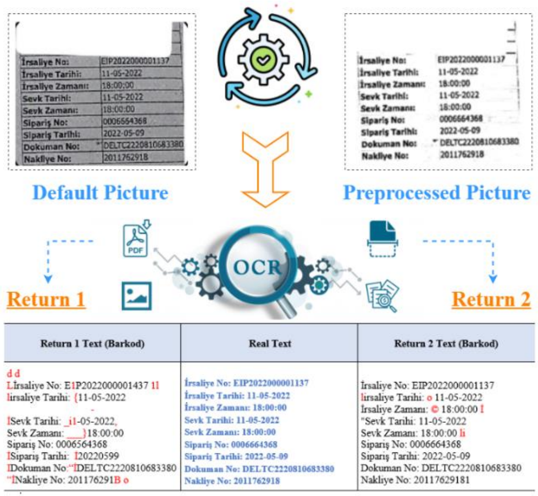


Figure 11. Image preprocessing for invoice information

When examining Figure 10 and Figure 11, it is observed that image preprocessing is very important to increase the efficiency of text recognition and processing software. Noise and distortions in the image prevent the OCR software from accurately recognizing the text. With the developed image preprocessing technique, noise has been effectively removed from many images. As a result, the edges have been enhanced, the contrast has been adjusted, the curvatures have been corrected, and different lighting conditions have been minimized to make the images more suitable for text recognition.

Furthermore, image preprocessing enhances the accuracy of OCR software by highlighting the differences between different font types. Particularly in the digitization of official documents using deep learning, image preprocessing is critical and increases the accuracy rate.

5. Conclusion and Discussion

Various steps such as machine learning, object detection, image processing, and OCR operations were performed in this study. As a result of these steps, some calculations that may vary were made. These calculations are visualized in Figure 12. The results obtained from the figure will be interpreted and evaluated in detail.

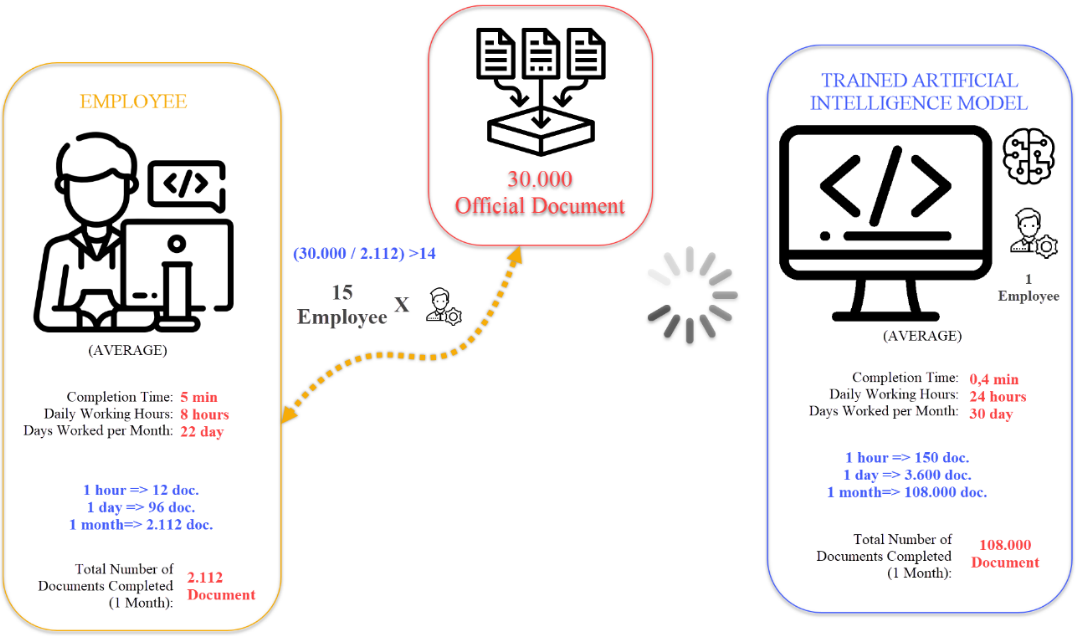


Figure 12. Comparison of workforce and cost between personnel and Artificial Intelligence model.

Assuming a monthly record of 30,000 official documents, based on the values shown in Figure 12, calculations are dependent on parameters such as completion time, daily working hours, and number of days worked per month, indicating that single personnel can process 2,112 documents per month. When the same parameters are applied to the AI model trained, a monthly average of 108,000 documents can be processed using the computer with the specifications listed in Table 3. This method results in clear time, cost, and workload savings.

The method developed in the study is superior to similar applications due to its accuracy, effectiveness, applicability, and presentation of alternative methods. Its difference from previous studies lies in the maximum interaction of deep learning, image preprocessing, object detection, text recognition, and processing methods with each other.

This study is believed to contribute to the development of artificial intelligence applications in many fields, especially in IT and R&D departments. As a result, a multidisciplinary working environment will be created. It is anticipated that this project will serve as a great example for digital transformation efforts in the industry and will be one of the leading projects.

Compliance with ethics requirements

This study does not contain any studies with human participants or animals performed by any of the authors.

Availability of Supporting Data

The data used to support the findings of this study are available from the corresponding author upon request.

Declaration of Competing 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.

Fundings

This publication is sponsored by ALISAN Logistics Inc.

Author Contributions Statement

Cuneyt ERGEN and Ensar GUNAYDIN developed the idea stage of the study. Bunyamin GENCTURK and Murat KOKLU collected the invoice data and performed data analysis using deep learning and text recognition processing techniques after preprocessing. Murat KOKLU supervised the project throughout. All authors contributed to the writing and editing of the manuscript.

Acknowledgements

In this project, ALISAN Logistics Inc. is a stakeholder that contributed to the emergence of the working idea and creation of data. We would like to express our gratitude to ALISAN Logistics Inc., which operates in Turkey and provides expertise and knowledge in the logistics industry.

References

[1] Zhao, G., et al. Skip-connected deep convolutional autoencoder for restoration of document images. in 2018 24th International Conference on Pattern Recognition (ICPR). 2018. IEEE.

[2] Subramani, N., et al., A survey of deep learning approaches for ocr and document understanding. arXiv preprint arXiv:2011.13534, 2020.

[3] Saha, R., A. Mondal, and C. Jawahar. Graphical object detection in document images. in 2019 International Conference on Document Analysis and Recognition (ICDAR). 2019. IEEE.

[4] Haralick, R.M. Document image understanding: Geometric and logical layout. in CVPR. 1994.

[5] Hashmi, K.A., et al., Current status and performance analysis of table recognition in document images with deep neural networks. IEEE Access, 2021. 9: p. 87663-87685.

[6] Soto, C. and S. Yoo. Visual detection with context for document layout analysis. in Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019.

[7] Li, K., et al. Cross-domain document object detection: Benchmark suite and method. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.

[8] Li, X.-H., F. Yin, and C.-L. Liu. Page object detection from pdf document images by deep structured prediction and supervised clustering. in 2018 24th International Conference on Pattern Recognition (ICPR). 2018. IEEE.

[9] Tran, D.N., et al., Table detection from document image using vertical arrangement of text blocks. International Journal of Contents, 2015. 11(4): p. 77-85.

[10] Qadri, M.T. and M. Asif. Automatic number plate recognition system for vehicle identification using optical character recognition. in 2009 International Conference on Education Technology and Computer. 2009. IEEE.

[11] Mori, S., H. Nishida, and H. Yamada, Optical character recognition. 1999: John Wiley & Sons, Inc.

[12] Chaudhuri, A., et al., Optical character recognition systems, in Optical Character Recognition Systems for Different Languages with Soft Computing. 2017, Springer. p. 9-41.

[13] Islam, N., Z. Islam, and N. Noor, A survey on optical character recognition system. arXiv preprint arXiv:1710.05703, 2017.

[14] Satyawan, W., et al. Citizen ID card detection using image processing and optical character recognition. in Journal of Physics: Conference Series. 2019. IOP Publishing.

[15] Le, A.D., D.V. Pham, and T.A. Nguyen. Deep learning approach for receipt recognition. in International Conference on Future Data and Security Engineering. 2019. Springer.

[16] Banerjee, J., A.M. Namboodiri, and C. Jawahar. Contextual restoration of severely degraded document images. in 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009. IEEE.

[17] Cai, J.-F., et al. Blind motion deblurring from a single image using sparse approximation. in 2009 IEEE Conference on Computer Vision and Pattern Recognition. 2009. IEEE.

[18] Bieniecki, W., S. Grabowski, and W. Rozenberg. Image preprocessing for improving ocr accuracy. in 2007 international conference on perspective technologies and methods in MEMS design. 2007. IEEE.

[19] Chen, X., et al. An effective document image deblurring algorithm. in CVPR 2011. 2011. IEEE.

[20] Bow, S.T., Pattern recognition and image preprocessing. 2002: CRC press.

[21] Cho, H., J. Wang, and S. Lee. Text image deblurring using text-specific properties. in European Conference on Computer Vision. 2012. Springer.

[22] Shan, Q., J. Jia, and A. Agarwala, High-quality motion deblurring from a single image. Acm transactions on graphics (tog), 2008. 27(3): p. 1-10.

[23] Tupaj, S., et al., Extracting tabular information from text files. EECS Department, Tufts University, Medford, USA, 1996. 1.

[24] Soni, R., B. Kumar, and S. Chand, Text detection and localization in natural scene images based on text awareness score. Applied Intelligence, 2019. 49(4): p. 1376-1405.

[25] Sun, W., et al., Deep-learning-based complex scene text detection algorithm for architectural images. Mathematics, 2022. 10(20): p. 3914.

[26] Patel, C., A. Patel, and D. Patel, Optical character recognition by open source OCR tool tesseract: A case study. International Journal of Computer Applications, 2012. 55(10): p. 50-56.

[27] Mursari, L.R. and A. Wibowo, The effectiveness of image preprocessing on digital handwritten scripts recognition with the implementation of OCR Tesseract. Computer Engineering and Applications Journal, 2021. 10(3): p. 177-186.

[28] Hegghammer, T., OCR with Tesseract, Amazon Textract, and Google Document AI: a benchmarking experiment. Journal of Computational Social Science, 2022. 5(1): p. 861-882.

[29] Matas, J., et al., Robust wide-baseline stereo from maximally stable extremal regions. Image and vision computing, 2004. 22(10): p. 761-767.

[30] Rezende, E., et al. Malicious software classification using transfer learning of resnet-50 deep neural network. in 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA). 2017. IEEE.

[31] Lowe, D.G., Distinctive image features from scale-invariant keypoints. International journal of computer vision, 2004. 60(2): p. 91-110.

[32] Liao, M., et al. Textboxes: A fast text detector with a single deep neural network. in Thirty-first AAAI conference on artificial intelligence. 2017.

[33] Ren, S., et al., Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 2015. 28.

[34] Al-Doori, S.K.S., Y.S. Taspinar, and M. Koklu, Distracted driving detection with machine learning methods by cnn based feature extraction. International Journal of Applied Mathematics Electronics and Computers, 2021. 9(4): p. 116-121.

[35] Gu, J., et al., Recent advances in convolutional neural networks. Pattern recognition, 2018. 77: p. 354-377.

[36] Koklu, M., I. Cinar, and Y.S. Taspinar, Classification of rice varieties with deep learning methods. Computers and electronics in agriculture, 2021. 187: p. 106285.

[37] Lin, P., et al., A deep convolutional neural network architecture for boosting image discrimination accuracy of rice species. Food and Bioprocess Technology, 2018. 11(4): p. 765-773.

[38] Krizhevsky, A., I. Sutskever, and G.E. Hinton, Imagenet classification with deep convolutional neural networks. Communications of the ACM, 2017. 60(6): p. 84-90.

[39] Yasin, E.T., I.A. Ozkan, and M. Koklu, Detection of fish freshness using artificial intelligence methods. European Food Research and Technology, 2023: p. 1-12.

[40] Unal, Y., et al., Application of pre-trained deep convolutional neural networks for coffee beans species detection. Food Analytical Methods, 2022. 15(12): p. 3232-3243.

[41] Koklu, M., H. Kahramanli, and N. Allahverdi, A new approach to classification rule extraction problem by the real value coding. International Journal of Innovative Computing,

Information and Control, 2012. 8(9): p. 6303-6315.
[42] Altun, A. and M. Koklu, Optimizing the learning process of multi-layer perceptrons using a hybrid algorithm based on MVO and SA. International Journal of Industrial Engineering Computations, 2022. 13(4): p. 617-640.
[43] Gencturk, B., et al., Detection of hazelnut varieties and development of mobile application with CNN data fusion feature reduction-based models. European Food Research and Technology, 2023: p. 1-14.
[44] Taspinar, Y.S., M. Koklu, and M. Altin, Fire Detection in Images Using Framework Based on Image Processing, Motion Detection and Convolutional Neural Network. International Journal of Intelligent Systems and Applications in Engineering, 2021. 9(4): p. 171-177.
[45] Butuner, R., et al., Classification of deep image features of lentil varieties with machine learning techniques. European Food Research and Technology, 2023. 249(5): p. 1303-1316.
[46] Koklu, M., et al. Identification of sheep breeds by CNN-based pre-trained InceptionV3 model. in 2022 11th Mediterranean Conference on Embedded Computing (MECO). 2022. IEEE.
[47] He, K., et al. Deep residual learning for image recognition. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[48] Simonyan, K. and A. Zisserman, Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[49] Wen, L., X. Li, and L. Gao, A transfer convolutional neural network for fault diagnosis based on ResNet-50. Neural Computing and Applications, 2020. 32: p. 6111-6124.
[50] Albawi, S., T.A. Mohammed, and S. Al-Zawi. Understanding of a convolutional neural network. in 2017 international conference on engineering and technology (ICET). 2017. Ieee.
[51] Hu, J., L. Shen, and G. Sun. Squeeze-and-excitation networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[52] Taspinar, Y.S., et al., Computer vision classification of dry beans (Phaseolus vulgaris L.) based on deep transfer learning techniques. European Food Research and Technology, 2022. 248(11): p. 2707-2725.
[53] Dogan, M., et al., Dry bean cultivars classification using deep cnn features and salp swarm algorithm based extreme learning machine. Computers and Electronics in Agriculture, 2023. 204: p. 107575.
[54] Kursun, R., et al. Flower recognition system with optimized features for deep features. in 2022 11th Mediterranean Conference on Embedded Computing (MECO). 2022. IEEE.
[55] Memon, J., et al., Handwritten optical character recognition (OCR): A comprehensive systematic literature review (SLR). IEEE Access, 2020. 8: p. 142642-142668.
[56] Mayya, V., et al., An empirical study of preprocessing techniques with convolutional neural networks for accurate detection of chronic ocular diseases using fundus images. Applied Intelligence, 2023. 53(2): p. 1548-1566.
[57] Rehman, A. and T. Saba, Neural networks for document image preprocessing: state of the art. Artificial Intelligence Review, 2014. 42: p. 253-273.
[58] Mittal, R. and A. Garg. Text extraction using OCR: a systematic review. in 2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA). 2020. IEEE.