RESEARCH ARTICLE

A novel deep learning method for recognizing texts printed with multiple different printing methods [version 1; peer review: awaiting peer review]

Jarmo Koponen, Keijo Haataja, Pekka Toivanen

School of Computing, Kuopio campus, University of Eastern Finland, Kuopio, Pohjois-Savo, FI-70211, Finland

v1 **First published:** 20 Apr 2023, **12**:427
https://doi.org/10.12688/f1000research.131775.1
Latest published: 20 Apr 2023, **12**:427
https://doi.org/10.12688/f1000research.131775.1

Abstract

Background: Text recognition of cardboard pharmaceutical packages with machine vision is a challenging task due to the different curvatures of packaging surfaces and different printing methods.
Methods: In this research, a novel deep learning method based on regions with convolutional neural networks (R-CNN) for recognizing binarized expiration dates and batch codes printed using different printing methods is proposed. The novel method recognizes the characters in the images without the need to extract handcrafted features. In detail, this approach performs text recognition considering the whole image as an input extracting and learning salient character features straight from packaging surface images.
Results: The expiration date and manufacturing batch codes of a real-life pharmaceutical packaging image set are recognized with 91.1% precision with a novel deep learning-based model, while Tesseract OCR text recognition performance with the same image set is 38.3%. The novel model outperformed Tesseract OCR also in tests evaluating recall, accuracy, and F-Measure performance. Furthermore, the novel model was evaluated in terms of multi-object recognition accuracy and the number of unrecognized characters, in order to achieve performance values comparable to existing multi-object recognition methods.
Conclusions: The results of this study reveal that the novel deep learning method outperforms the well-established optical character recognition method in recognizing texts printed using different printing methods. The novel method presented in the study recognizes texts printed with different printing methods with high precision. The novel deep learning method is suitable for recognizing texts printed on curved surfaces with proper preprocessing. The problem investigated in the study differs from previous research in the field, focusing on the recognition of texts printed with different printing methods. The research thus fills a gap in text recognition that existed in the research of the field. Furthermore, the study presents

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

new ideas that will be utilized in our future research.

Keywords
Text recognition, Machine Vision, Deep Learning, Regions with Convolutional Neural Networks, R-CNN, Character Recognition, Printing Methods, Expiration Date, Batch Codes, Handcrafted Features, Image Recognition, Multi-Object Recognition, OCR, Tesseract OCR, Precision, Recall, Accuracy, F-Measure, Curved Surfaces, Preprocessing

This article is included in the Artificial Intelligence and Machine Learning gateway.

Corresponding author: Jarmo Koponen (jarmo.koponen@uef.fi)

Author roles: Koponen J: Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing; **Haataja K**: Funding Acquisition, Project Administration; **Toivanen P**: Supervision, Validation

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

How to cite this article: Koponen J, Haataja K and Toivanen P. **A novel deep learning method for recognizing texts printed with multiple different printing methods [version 1; peer review: awaiting peer review]** F1000Research 2023, **12**:427
https://doi.org/10.12688/f1000research.131775.1

First published: 20 Apr 2023, **12**:427 https://doi.org/10.12688/f1000research.131775.1

Introduction

By the end of 2023, the market value of the pharmaceutical packaging sector is expected to reach USD 101 billion.[1] Recognizing product codes using machine vision enables the storage and processing of package-specific manufacturing data, as well as the electronic search and extraction of codes and dates, which is important in the development of intelligent product handling systems.

Cardboard is typically used in pharmaceutical product packaging. Boxes made of cardboard have curved surfaces. When accurate text recognition of text printed on surfaces is needed using 2D machine vision, the curvature is a difficulty in itself as it causes uneven illumination of the packaging surface as presented in Figure 1. Several printing methods are used to print the expiration date and manufacturing batch codes that are important for the usage and handling of pharmaceutical packages. Changes in physical conditions in package imaging as well as printing methods that generate low-contrast text and irregular letter forms make it more difficult to recognize these codes. The recognition process is especially challenging because there are many variants in the codes used on various packages using different printing methods, as well as variances in the codes' forms, structures, regularity, and colors. In the past ten years, product code recognition techniques have advanced, and several researchers have become more interested in this area of study.

The purpose of this research is to demonstrate that, despite field difficulties, texts with imperfections printed on pharmaceutical packaging can be effectively recognized using appropriate pre-processing and deep learning. In the experimental part of the study, the recognition accuracy of the text recognition method trained on a real-life image set and the number of unrecognized characters is measured using generally comparable metrics, and the novel deep learning text recognition method is compared to the Tesseract OCR method in the recognition of dot matrix and ink-printed characters using four multi-object recognition metrics. It is important to note that information related to expiration dates and manufacturing batch codes on pharmaceutical packages is important for safety and must be accurate and reliable. The rest of the paper is organized as follows: we discusses the related works from a survey of the literature. A description of the proposed methodology is then presented. Analysis of experimental results is then presented and finally, the conclusion and future works are presented in the final section.

Related works

The product's manufacturing markings are recognized by a computer and optical character recognition software from the digitized image produced by an imaging system, which is based on the amount of light energy reflected from the surfaces of the objects in the scene.[2] Optical Character Recognition (OCR) methods that compared groups of pixels detected on the surface of a product to models given to the system have been increasingly overtaken by methods utilizing deep learning technology.[2] OCR software can recognize characters that are regularly shaped and have good contrast against a simple background.[3] However, the difficulties in recognizing packaging texts, which are described with solutions,[2] reduce the usability of the OCR method in this high-accuracy task.

Tesseract OCR is a widely used open-source optical character recognition program. It has been in use since 1985 and has evolved significantly throughout the years. Over time, Tesseract OCR has changed. Numerous languages are supported, image binarization is included, and Tesseract 4 added an OCR engine with an LSTM-based neural network, which

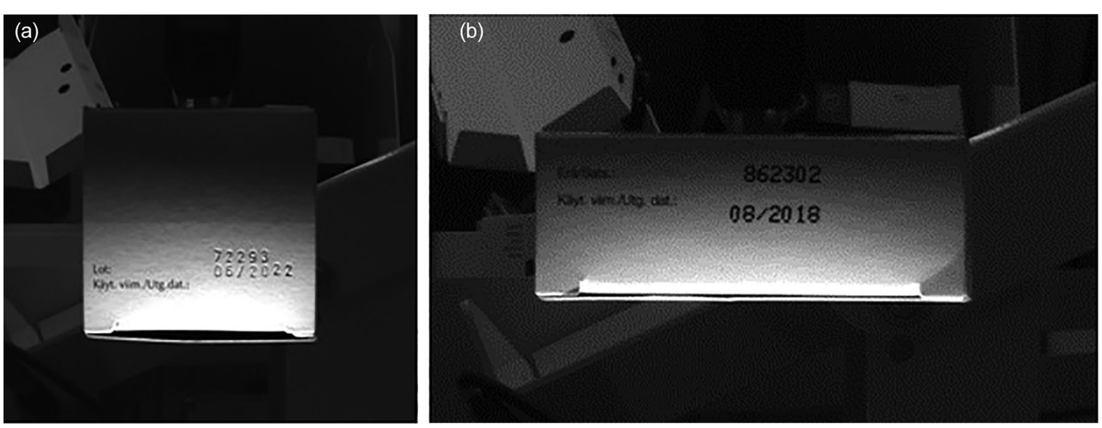

Figure 1. Two cardboard pharmaceutical packages with the expiration date and batch codes printed on the curved packaging surfaces.

process an input image line by line into boxes and then feeds them to the LSTM network, which produces the recognition result.[4]

OCR text recognition was used in the research by Gong *et al.* (2020) to recognize text in images of food packaging. The text that was printed on the product package had slanted characters. Some images were light-exposed, and some texts were printed with poor print quality. The Tesseract OCR method correctly recognizes 31.1% of characters.[5] In the same study, the deep learning text recognition method achieves 95.4% recognition accuracy with identical data.

Deep neural networks are being used more frequently for product text recognition as of 2018, which has made it possible to recognize inconsistent characters, detect and recognize the text of various sizes in images taken in real-world conditions, and even recognize texts from moving packaging.[2] Before that, it appears that conventional recognition methods were more frequently employed.[2]

Methods

This research focused on utilizing deep learning to recognize texts printed on the surface of differently curved packages. Dot matrix printed texts and pressed ink-marked texts are the focus of our study. As part of our research work, we have developed a novel algorithm for recognizing text from images taken from packages. The source images are taken in a controlled imaging environment and binarized with a method in Ref. 6, which, despite the variable contrast between the text and the background and the curvature of the packaging surface, is capable of accurate and robust binarization. The novel method's text recognition ability is first tested in terms of text recognition accuracy and the number of unrecognized characters. Furthermore, using four criteria, the performance of the deep learning network constructed for text recognition is compared against Tesseract OCR. Text recognition of texts pressed without ink, stamped, or laser printed is excluded from the study. In the experimental recognition tests, we used a set of images of a Finnish health technology company taken from real medical packages, of which 16 pcs were with dot matrix printed texts, and 11 pcs with pressed ink-marked texts. The study's text recognition was performed on the packaging production batch and expiration date markings, and titles. An example of the binarized images used for text recognition is shown in Figure 2. In the experimental tests, real pharmaceutical packages have been used to ensure that the results of our research are reliable. Since the text recognition of pharmaceutical packages is a new area of research, the number of source images was limited. However, this limitation is only temporary, and research will continue as more source images become available.

Deep Learning

Deep Learning (DL) is a subset of machine learning in which neural networks are used to learn from data and generate predictions or choices. It is described as a method of teaching a computer to learn from data and make predictions or judgments using neural networks, much like a child does. Deep Learning can be used for a variety of tasks, including image classification, speech recognition, and natural language processing.

R-CNN for expiration date and batch codes recognition

Regions with Convolutional Neural Networks (R-CNN) architecture combines rectangular region proposals with convolutional neural network features. Figure 3 represents the R-CNN architecture in character recognition. For character recognition, R-CNN first extracts several object-independent region proposals from the source image using the method of selective search (Figure 3a). Each region is sent on to the convolution neural network, for the creation of a fixed-size

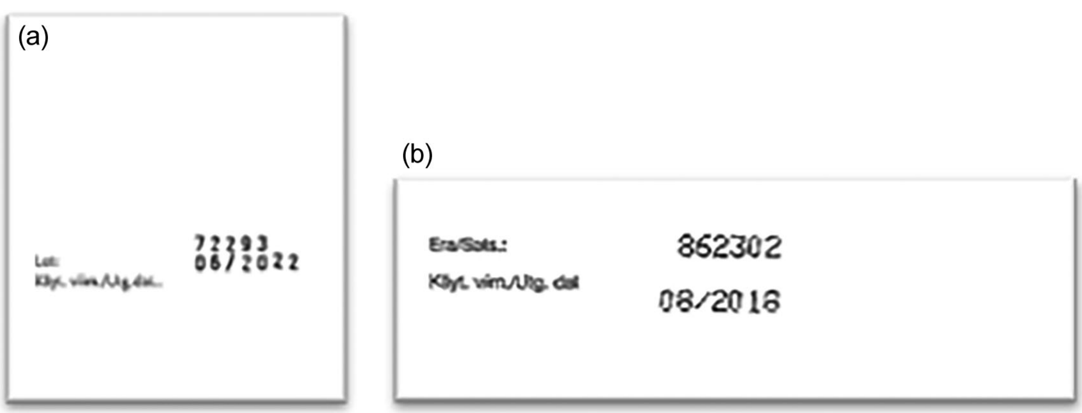

Figure 2. Batch code and expiration date texts in a binarized images printed with a pressed ink-marked (left) and with the dot matrix printing methods.

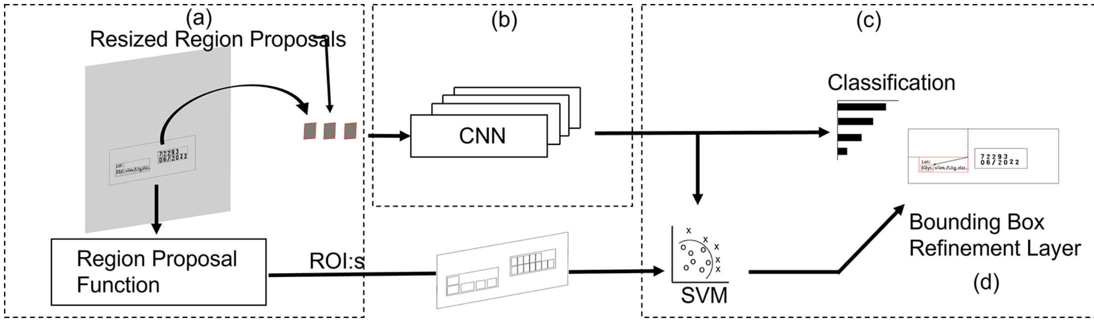

Figure 3. Region-CNN model working principle.

feature vector corresponding to them (Figure 3b). The feature vector is then used as the input to a set of linear SVMs that are trained for each class and output a classification (Figure 3c). To obtain the most accurate coordinates, the vector is also fed into a bounding box refinement layer of a network (Figure 3d).[7,8]

Training details

The dataset contains 27 binarized source images, 16 of which have dot matrix texts and 11 of which have pressed, and ink-marked texts. The network was trained using the stochastic gradient descent with a momentum optimization algorithm, with a learning rate of 1e-4 that is reduced by 0.02 after every eighth epoch. Other network training parameters include max epoch 400, and a mini-batch size 58. A pre-trained Alex-Net network with transfer learning is combined with a Region-CNN deep learning network to recognize the 43 different categories of objects in the images. First, images are fed into Alex Net CNN (i.e., $227 \times 227 \times 3$). The dataset is divided into two parts: training and validation. The final layers of the pre-trained Alex Net CNN were modified to match the number of classes in the training dataset. To prevent network overfitting, data augmentation was used for training images. Modified Region-CNN was trained using the augmented training images and labels.

Results

The experiments aimed to evaluate the recognition accuracy of the deep learning method for text recognition on real-world pharmaceutical packages. The experiments were carried out on available dot matrix, printed, and ink-marked texts. The recognition results were achieved by comparing the recognition results obtained with the test image set to the ground truth data.

R-CNN evaluation in MATLAB

The accuracy of the text recognition of the novel model was analyzed in the first test with a dataset using a generally comparable evaluate Detection Precision[9] function. The function outputs the accuracy of each object class for a multi-object detector. Also, the evaluate Detection Miss rate[10] function was used in the evaluation, calculating the model's performance based on the number of undetected objects in the recognition of multiple objects. Figure 4 shows the mean Average Precision- and log Average Miss Rate-curves plotted for all object categories in the character recognition test data set.

Recall and precision are cell arrays for a multi-object detector, with each cell containing data points for each object category. Log Average Miss Rate (LAMR) is also calculated for each object category separately. The mean Average Precision (mAP) metric was used to evaluate the performance of the multi-object detection model. The mAP is calculated by taking the mean of the AP values for each object category yielding a Precision value of 0.96 and a Recall value of 0.72. The detector's performance findings demonstrate the percentages of accurate classification (precision), and all searchable objects detected (recall). The mAP rating of 0.72 indicates that the detector performed well, produced accurate findings, and found a large percentage of the objects analyzed. The Log Average Miss Rate (LAMR) function was used to analyze the performance of the multi-object recognition task. The measured LAMR value of 0.13 demonstrates that the number of undetected objects was very low, and that the detector performed well in the multi-object detection test.

Comparison to Tesseract OCR

In the second experiment, the same data set with images containing text was given for text recognition to both the novel deep learning model and Tesseract OCR. The numbers of all characters for each image were first counted. The number of correctly identified and incorrectly recognized characters, as well as the number of unrecognized and double-identified characters, are calculated for performance evaluation. This gave detailed information on the methods' text recognition accuracy. The threshold value (α) of the R-CNN model in the recognition accuracy test was set to 0.99.

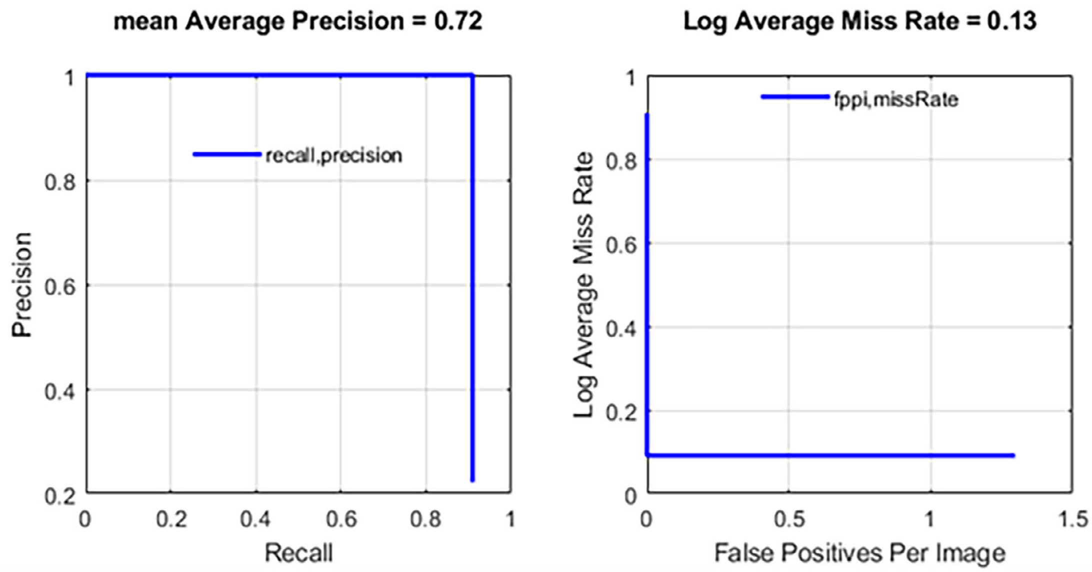

Figure 4. The proposed text recognition model's mean average precision and log average miss rate analysis results.

IoU (Intersection over Union) is a common object detection evaluation metric. IoU is a measure of the intersection being the overlap between the ground truth boundary box and the algorithm-detected boundary box.[11] IoU, as depicted in Equation (1), is calculated as the area of overlap between the ground truth bounding box and the predicted bounding box divided by the area of the union of these two.[11]

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\text{area (ground truth bounding box} \cap \text{predicted bounding box)}}{\text{area (ground truth bounding box} \cup \text{predicted bounding box)}} \quad (1)$$

Introducing Key Metrics for Object Detection: TP, FN, TN, FP, Precision, Recall, Accuracy, and F-Measure

Positive and negative samples are required for all supervised learning methods. Using a human face detector as an example, images with faces are positive samples, whereas images without faces are negative.[12] After testing, a group of positive and negative samples can be classified into one of four states, as shown in Table 1.[12]

Positive samples may create the following scenarios:

1. TP (true positive), in which a positive sample is determined as the target by the detector, IoU $\geq \alpha$.

2. FN (false negative), in which a positive sample is determined by the detector to be non-target.

Negative samples can create the following scenarios:

3. TN (true negative), in which a negative sample is determined by the detector to be non-target.

4. The detector determines FP (false positive), or a negative sample, as the target, IoU $< \alpha$.

Table 1. Prediction result of the sample.

	Actual positive	Actual negative
Predicted positive	True positives (TP)	False positives (FP)
Predicted negative	False negatives (FN)	True negatives (TN)

These formulas are used to calculate precision and recall:

Precision is the proportion of all correctly retrieved samples (TP) that account for all retrieved samples (TP + FP).[12] Its equation is, as given in (2):

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{2}$$

The recall is the proportion of all correctly retrieved samples (TP) that accounts for all objects that should be retrieved (TP + FN).[12] Its equation is, as given in (3):

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3}$$

Accuracy is the proportion of all correctly retrieved samples that account for all samples.[12] Its equation is, as given in (4) as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{4}$$

The F-Measure, as depicted in Eq (5), is the weighted harmonic average of precision and recall.[12]

$$\text{F-Measure} = \frac{2*(\text{Precision}*\text{Recall})}{\text{Precision} + \text{Recall}} \tag{5}$$

Table 2. Text recognition performance comparison table based on four metrics.

	Region-CNN (%)	Tesseract OCR (%)
Precision	91.1	38.3
Recall	72.7	61.3
Accuracy	69.9	30.8
F-Measure	80.9	47.1

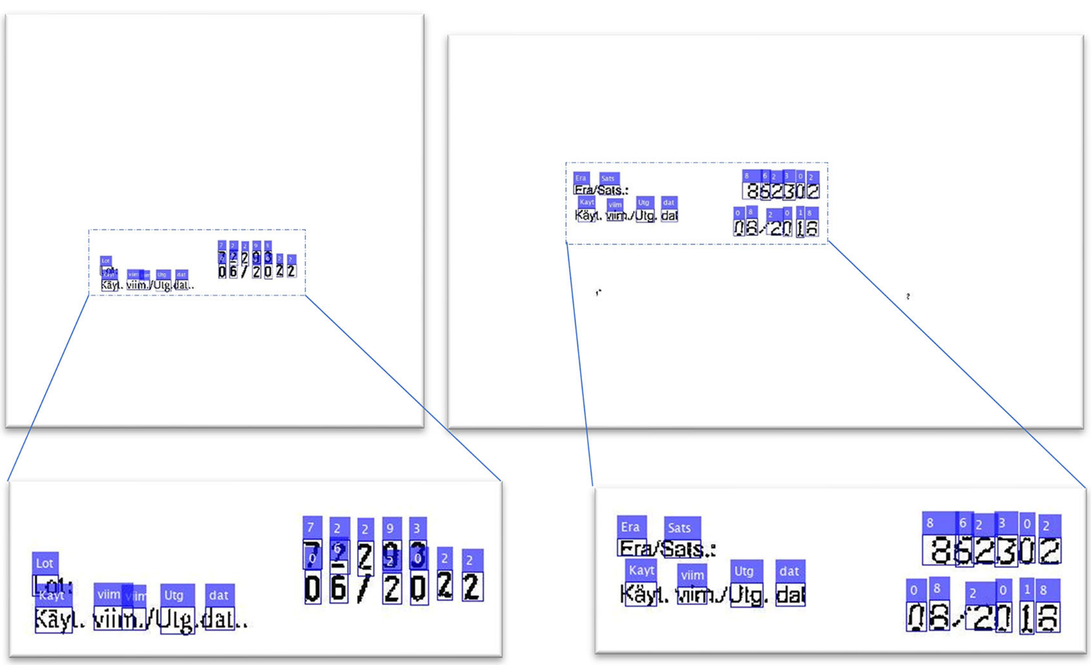

Figure 5. The result of deep text recognition in the whole image. Left: recognized printed ink-marked packaging text. Right: recognized dot matrix printed packaging text.

Using these four different metrics the text recognition performance of the novel deep learning and Tesseract OCR algorithms are compared in Table 2.

The comparison test results demonstrate that the developed deep learning character recognition accuracy outperforms the OCR method significantly. The deep learning model had a recognition precision of 91.1%, while the OCR method had a recognition precision of 38.3%. The novel deep learning model was significantly more precise and efficient than the OCR method in recognizing the test set's texts. The deep learning model has a recall value of 72.7%, while the Tesseract method has a recall value of 61.3%. When compared to the Tesseract OCR approach, the novel model's recall value in text recognition of the target test set is 11,4% higher. On the test image set, the deep learning model's text recognition accuracy was 69.9%, whereas the Tesseract method's text recognition accuracy was 30,8%. The F-Measure is a number that measures the object detector's precision-to-recall ratio. Precision describes the proportion of correct recognitions made out of all retrieved recognitions, whereas recall refers to the proportion of correct recognitions accounts for all recognition's should have been retrieved. The F-Measure combines these two results into a numerical result that describes the object detector's overall performance. The F-Measure of the novel deep learning model is 80.9%, which is higher than the F value of the Tesseract OCR text recognition method, which is 47.1%. This shows that the deep learning model outperforms in terms of both Precision and Recall. The results demonstrate that the novel deep learning model is a more efficient text recognition method and can recognize characters in test images more consistently. The text recognition results using the novel Region-CNN method are shown in Figure 5.

Conclusions

In this paper, deep learning-based text recognition of pharmaceutical packages has been studied, and our novel method for expiration date and batch codes text recognition was presented. The recognition results achieved by a novel method, which is based on a regional convolution neural network, appear to be excellent and significantly outperform those obtained by Tesseract OCR, allowing the recognition of texts with inconsistencies in character shapes printed on curved surfaces with proper preprocessing.

Various methods for recognizing text printed on products have been developed. Commercial OCR engines are available that can recognize high-contrast regularly shaped texts printed on flat surfaces. In the industry, large amounts of perishable food and pharmaceutical packages are produced daily, and the texts printed on their surface is read by people several times during the operation. Text is printed on product surfaces using various printing methods to achieve cost-effective production. The pharmaceutical industry employs various printing methods, such as laser printing, pressed without ink, and pressed with ink, resulting consistent character shapes. However, characters printed using dot matrix and manual stamping methods may have inconsistencies in their character shapes. In places where pharmaceutical product packaging is handled, such as pharmacies and drug storage facilities, where it is essential to recognize texts printed with various methods, a targeted method is required.

We compare the text recognition performance obtained using a Tesseract OCR and our novel targeted method with real-life pharmaceutical packaging in this study. We find that the novel deep learning method produces highly accurate recognition results that completely outperform the results obtained using a Tesseract OCR, indicating the limitations of optical text recognition for the research's text recognition needs.

Although text recognition for pharmaceutical packaging texts printed with different methods is an essential task, it is still in its early stages due to the limited availability of source images of actual pharmaceutical packages for research purposes.

To the best of our knowledge, this is the first study in which deep learning is used to recognize inconsistencies in texts printed using different printing methods. We are confident that with more training data, the proposed method's performance would increase even more.

In the following study, we focus on the domain-specific contextual processing of recognized text characters. This results in a three-phase text recognition pipeline targeted at recognizing manufacturing markings on pharmaceutical packaging. With this approach, we improve the understanding of how deep learning can be applied in this field.

Data availability

Underlying data

Open Scientific Framework: The underlying data for A novel deep learning method for recognizing texts printed with multiple different printing methods, https://doi.org/10.17605/OSF.IO/MP3RB.[13]

This project contains the following underlying data:

- Novel text recognition methods source code for the Octave software.
- RCNN and OCR recognition methods metrics spreadsheet.xlsx (An Excel spreadsheet containing the achieved recognition results).

References

1. Global Packaging Market Size. 2020.
Reference Source

2. Koponen J, Haataja K, Toivanen P: **Recent advancements in machine vision methods for product code recognition: A systematic review.** *F1000Res.* 2022; **11**(1099): 1099.
Publisher Full Text

3. Althobaiti H, Lu C: **A survey on Arabic optical character recognition and an isolated handwritten Arabic character recognition algorithm using encoded freeman chain code.** *Paper presented at the 2017 51st Annual Conference on Information Sciences and Systems, CISS 2017.* 2017.

4. **Tesseract User Manual | tessdoc (tesseract-ocr.github.io).**

5. Gong L, Thota M, Yu M, *et al.*: **A novel unified deep neural networks methodology for use by date recognition in retail food package image.** *SIViP.* 2020; **15**(3): 449–457.
Publisher Full Text

6. Koponen J, Haataja K, Toivanen P: **Text Recognition of Cardboard Pharmaceutical Packages by Utilizing Machine Vision.** *Electronic Imaging.* 2021; **33**: 235-1–235-7.
Publisher Full Text

7. Wang Y, Liu X, Tang Z: **An R-CNN based method to localize speech balloons in comics.** *MultiMedia Modeling: 22nd International Conference, MMM 2016, Miami, FL, USA, January 4-6, 2016, Proceedings, Part I 22.* 2016; pp. 444–453.

8. Girshick R, Donahue J, Darrell T, *et al.*: **Rich feature hierarchies for accurate object detection and semantic segmentation.** *Proceedings of the 2014 Conference on Computer Vision and Pattern Recognition.* 2014.

9. Mathworks Inc: **Precision Detection.** Accessed 30 January 2023.
Reference Source

10. Mathworks Inc: **MissRate Detection.** Accessed 30 November 2023.
Reference Source

11. Nachappa CH, Rani NS, Pati PB, *et al.*: **Adaptive dewarping of severely warped camera-captured document images based on document map generation.** *Int. J. Doc. Anal. Recognit.* 2023; 1–21.
PubMed Abstract | Publisher Full Text | Free Full Text

12. Gong S, Liu C, Ji Y, *et al.*: *Advanced image and video processing using MATLAB.* Vol. 12. Springer; 2018; p. 581.

13. Koponen J: **Recognition results of a Novel deep learning method for recognizing texts printed with multiple different printing methods.** 13 March 2023.
Publisher Full Text