International Journal on
"Technical and Physical Problems of Engineering"
(IJTPE)
Published by International Organization of IOTPE

ISSN 2077-3528
IJTPE Journal
www.iotpe.com
ijtpe@iotpe.com

December 2022 | Issue 53 | Volume 14 | Number 4 | Pages 363-369

DYNAMIC VISUALIZATION OF AN IMAGE FOR INTERACTIVE ACTIONS

V. Jain [1] Y. Jain [1] H. Dhingra [1] A. Jain [1] M. Demirci [2] M.C. Taplamacioglu [2]

1. Information Technology Department, Bharati Vidyapeeth's College of Engineering, New Delhi, Delhi, India
vanita.jain@bharatividyapeeth.edu, yugantar8jain@gmail.com, hardikdhingra@gmail.com, achin.mails@gmail.com
2. Electrical and Electronics Engineering Department, Gazi University, Ankara, Turkey
mervedemirci@gazi.edu.tr, taplam@gazi.edu.tr

Abstract- An image is generally static which is readable by the users but not actionable. On the other hand, a file type such as a PDF document is actionable with support for hyperlinks for phone numbers, email addresses, and more. In this paper, we propose the technology that can dynamically visualize an image and make it directly actionable for easy access to phone numbers, email addresses, and web links. The method uses state-of-the-art technologies, including Apple Developer's Vision Framework for text extraction and localization, NSDataDetector API for text classification, and open API for deep linking. We have conducted multiple tests and compared our method in terms of speed, overall accuracy, and data type-specific accuracy with different papers on multiple datasets. The experimental results from the present study test show that the proposed method gives the top of line performance in all contexts for speed and accuracy.

Keywords: Actionable, Data Classification, Dynamic Image, Text Extraction.

1. INTRODUCTION

Digital images are used abundantly around the world for different purposes. They are used to capture moments and share information, capture text, distribute posters and publicity materials, send invites, business cards and more. These static images often contain textual data that can be of use to the users (for example, phone numbers, email addresses, and web links). Some cases QR Codes also includes similar information inside [1, 2]. This paper proposes a method to dynamically visualize a static image and enable users to directly click on the data points such as phone numbers, email addresses, and web links on the image. This shall act as a great utility and significantly increase the convenience of acting on the data contained in an image, particularly helpful in the case of posters, business cards, publicity images, and other information-based images. To make a robust visualizer for dynamic images, the Apple Developer's Vision framework [3] is used that provides top-of-the-line accuracy for text extraction along with swift processing time.

The dynamic visualization is a four-step process, as shown in Figure 1. It involves:
1) Text extraction from image
2) Classification and localization of extracted text
3) Actionable button overlay on the frame of text strings
4) Platform-specific actions

The text extraction from the static image, we have used the Apple Developer Vision framework is used to detect and localize text from the image with extremely high speed and accuracy. After extracting the text from the image, the data into standard text, phone numbers, email addresses, and web links using the NSDataDetector API [4] are classified. An actionable button is placed as an overlay over the actionable text bounding box regions, and this is done using the coordinates of its frame/bounds detected in the first step through localization. When a user presses this button, a specific action (platform dependent) for the particular data type is triggered. In this work developed actions for phone numbers, email addresses, and web links; these are initiating a phone call to the respective number, composing a mail in the mail app to the respective email address, and opening the website in the browser on the iOS platform are implemented.

2. RELATED WORK

An extensive study is conducted on this, and different methods and algorithms are analyzed and then compiled. X. Zhou, et al. [5] use the Efficient and Accurate Scene Text Detector (EAST) algorithm to analyze the letters and words and convert them into machine-readable form. Zhang, et al. [6] talk about using a trained Fully Convolutional Network (FCN) model for text detection. B. Shi, et al. [7, 8], in their two papers, use Convolutional Recurrent Neural Network (CRNN), Sequence Recognition Network (SRN) and Robust Text Recognizer with Automatic Rectification (RARE) with Spatial Transformer Network (STN) and respectively for text recognition. Z. Cheng, et al. [9] use Arbitrary Orientation Network (AON) for the same purpose. Lee, et al. [10] are proposed the use of Recursive Recurrent Neural Networks with Attention Modelling (R22AM). M. Jaderberg, et al. [11] are used CNN with CRF for text extraction.

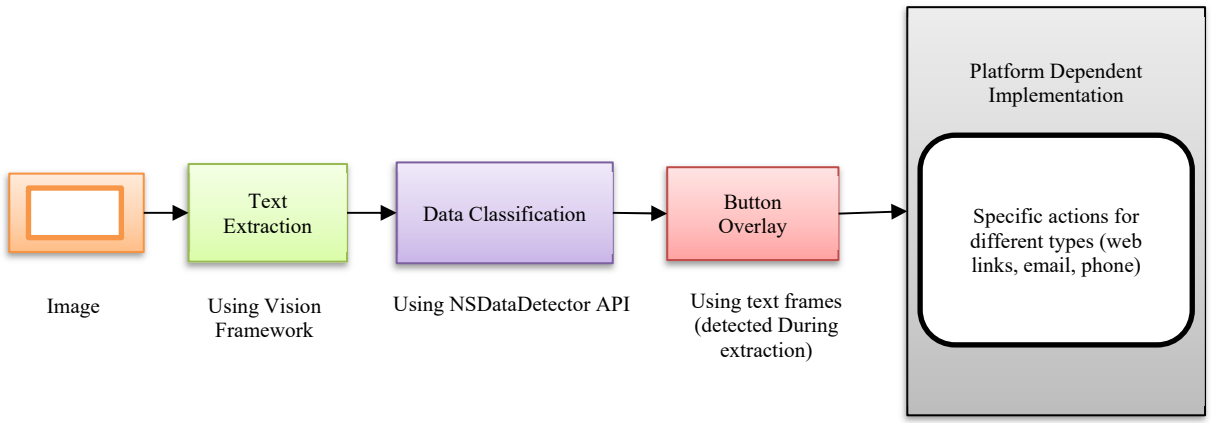


Figure 1. Flow diagram for dynamic image visualization

C. Madan Kumar, et al. [12] have developed algorithms for classifying text into a wide range of data types such as web links, email addresses, etc. The authors use keyword checking as an initial check for a lot of classes. For example: checking the "www" prefix for web links, "@" for email addresses, numerical digits for phone numbers and so on. They have used the package "libpostal" for parsing of string for the above classifications and Tesseract OCR for text extraction. The authors are applied text classification for 15 different data points and the work in this paper achieves greater than 95% accuracy for each class. In the paper, Alper Kursat Uysal and Serkan Gunal [13] have evaluated dimension reduction, text domain, classification accuracy and text language. All preprocessing tasks for text classification were examined in two different languages (e-mail and news) using two languages, Turkish and English, in this study. They have performed an experimental analysis on the datasets to show combinations of preprocessing tasks that might be suitable for achieving a major improvement in classification accuracy.

3. PROPOSED METHODOLOGY

As mentioned earlier, the dynamic visualization of a static image happens in a series of steps. The detailed working is described as follows.

3.1. Text Extraction

The present paper study offers two different paths for text extraction, Accurate and Fast behave as implied. While the correct path is better suited for application, we have presented the results for both the paths (fast and accurate) in the results section.

Our text extraction flow works by first getting the static image and using its data to build a Core Graphics [14] image object. Using this Core Graphic image object, it is initiated as VNImageRequestHandler which is an object that handles image processing results. In some cases where the image is rotated, it can be specified an additional parameter for orientation in our request handler object. Next, it is defined as the actual request for text extraction from the image using an VNRecognizeTextRequest object.

This object attaches our function for text recognition and further processing to itself through its handler. This function for text recognition and further processing includes recognizing and localizing the array of text strings from the image, retrieving bounding boxes, and placing clear buttons at the bounding frames. In this function, the results for our text extraction request are obtained as an array of VNRecognizedTextObservation objects; these observations contain the data of the text strings as well as the bounding box frames and are hence used for the extraction and processing of image. Finally, we perform the request using our request handler object and the request object we created earlier. This whole flow chart is shown visually in Figure 2.

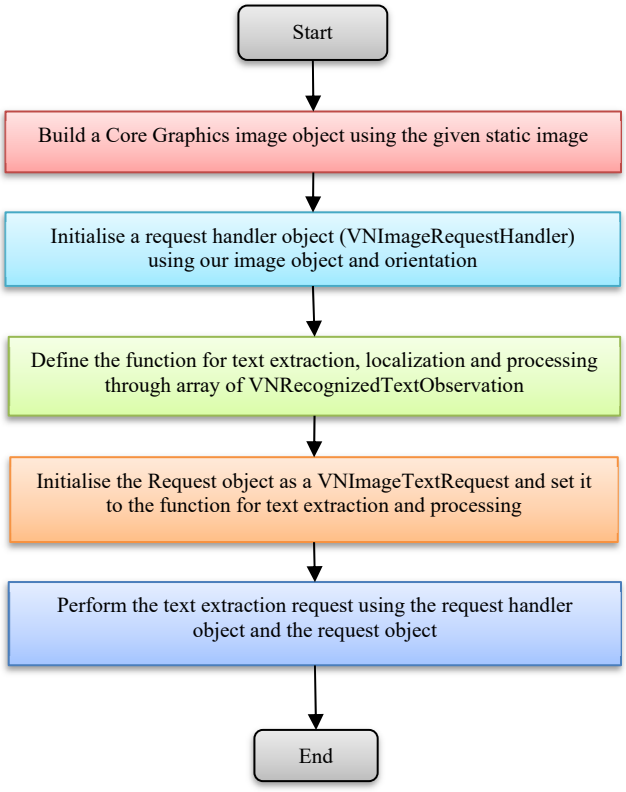


Figure 2. Text Extraction and Processing

3.2. Data Classification

To classify the text extracted from the image, we use the Apple Developer NSDataDetector API [4] is used. It is used to classify the data into the following types (which then form the basis for specialized actions):

- Phone Numbers
- Email Addresses
- Web Links

This API has helped quickly and accurately classify all the data points from the used images with text (business cards) into suitable data types. To detect the data type of the text extracted from the image, first an NSDataDetector object while specifying the types to be checked for is created. Following, detector object to get matches for the data types mentioned in the data is used. We get a match's sequence and each match contains data and meta-data (through properties) about the type. Next, we switch through the different types of data that it can have and define the action to be performed for each as described below.

3.3. Actions for Specific Data Types

Specific actions are triggered for different data types when users interact with the static image by tapping an actionable text area (phone number, email addresses, and/or web links). Table 1 gives a comprehensive overlook of the data types handled and the actions for them. It shall be noted that the actions are platform dependent as well as their implementation. We have used the iOS platform for our implementation. To open the browser, the mail app and calling on the user device, deep linking technology is used. For that we have used the "open (:options:completionHandler:)" API of UIKit [15].

Table 1. Data types with their specific actions

Data Type	Action Performed
Phone Number	Call initiated
Email Address	Mail app opened with pre-filled email address
Web Link	Link opened in Safari browser

4. RESULTS

Firstly, the result of the application and how it works below through an example of an image of a business card containing all the three data types: phone number, email address, and a web link are presented. When a phone number is tapped on an image, the call is automatically initiated with the phone number as shown in Figure 3.

When an email address is tapped on an image, a compose email action is initiated with the default email app with the receiver's address pre-filled automatically as shown in Figure 4. When a web link is tapped on an image, website is opened in default browser as shown in Figure 5.

To test our work for dynamic visualization of a static image for interactive actions on its text, we have used the' Stanford Mobile Visual Search Data Set: Business Cards' dataset [16] by V. Chandrasekhar, et al. This dataset contains a lot of real-life images of business cards containing actionable textual data including phone numbers, email addresses, and/or web links. Many of the images contain multiple points of data; this pushes technology to its limits and checks for real life scenarios including not only the actual text but also the localization capabilities. The Vision framework has helped the research to achieve extremely high performance and accuracy. It features two paths for text extraction, fast and accurate, and it is tested as application through the dataset for both of these paths. The speed comparison between the two paths is shown in Figure 6.

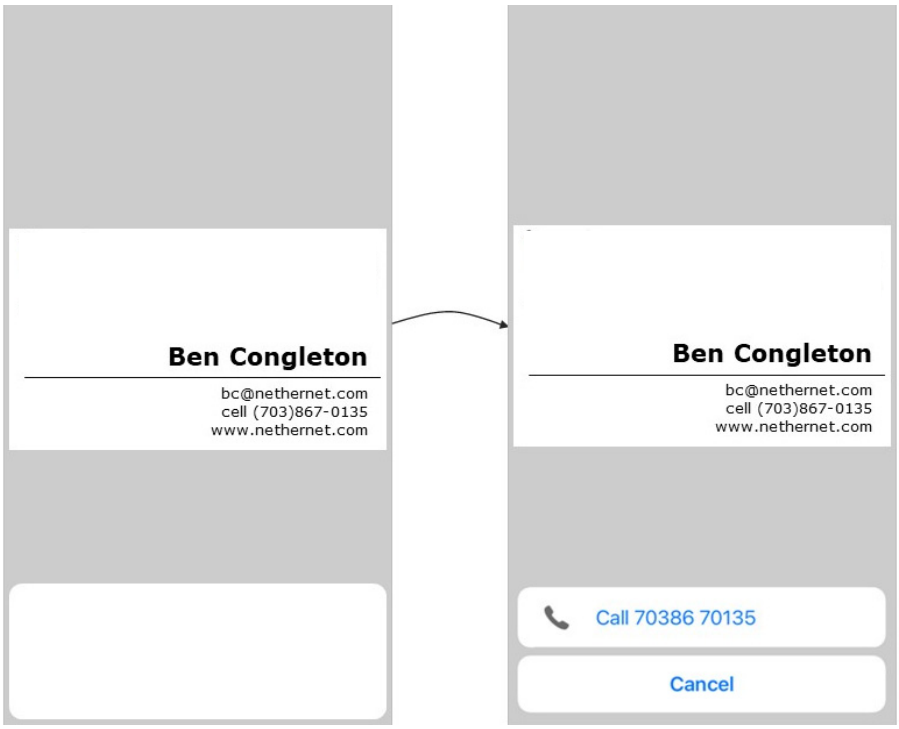


Figure 3. Call action

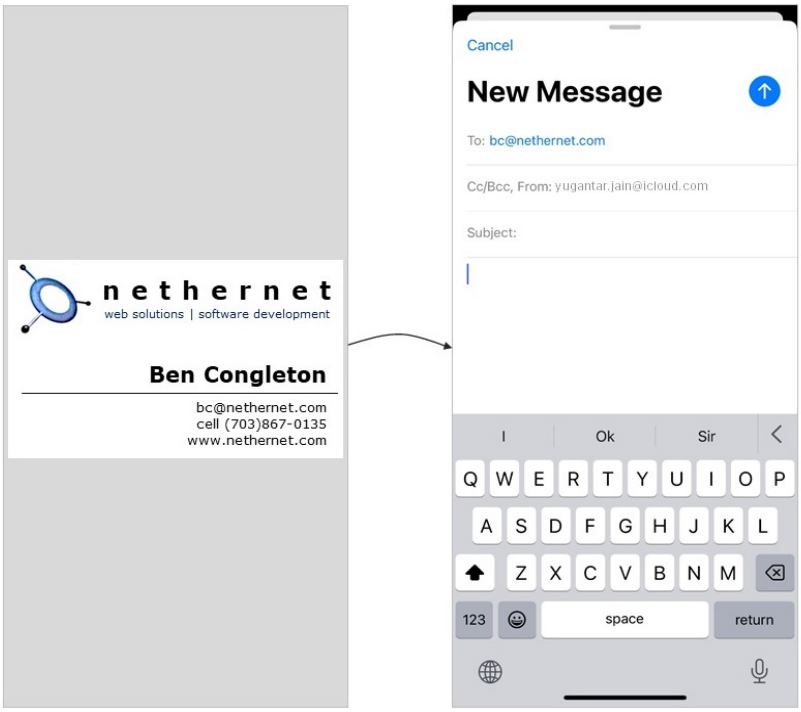


Figure 4. Email action

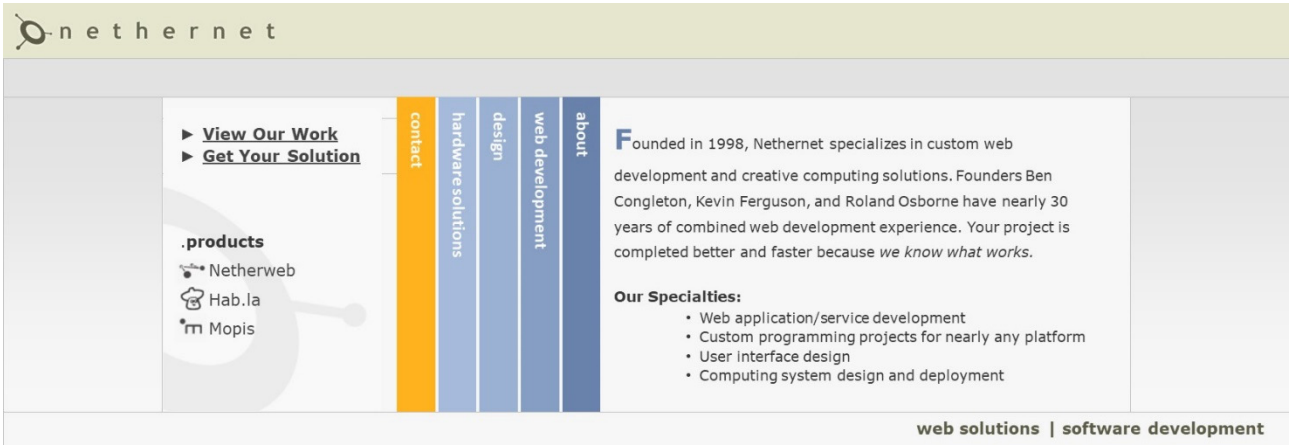


Figure 5. Web action

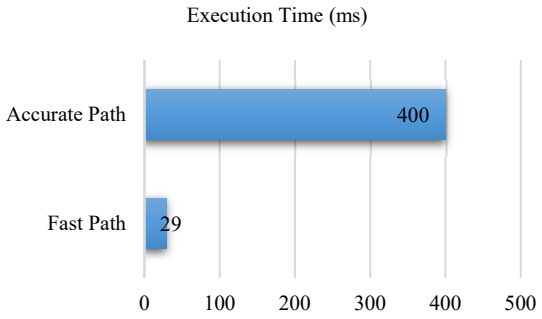


Figure 6. Performance results of fast and accurate paths in vision for text extraction

The system used for testing is packed with the ARM-based Apple M1 chip [17] containing 8 CPU cores with max clock rate of 3.2 GHz. While the accurate path is quite fast with an average text extraction speed of 0.4 seconds per image, the fast path is even faster (14 times faster) and takes around 0.029 seconds per image. However, the fast path is based on traditional OCR whereas the accurate path is built on sophisticated neural networks which makes the accurate path much more reliable. The accuracy comparison of both the paths against Madan Kumar, et al. [12] for correct text extraction of phone numbers, email addresses, and web links is shown in Figure 7. It is evident that the accurate path is much more reliable and accurate than its counterparts, with an average overall accuracy of 98.8%. Along with top of the chart accuracy, this path also gives a very fast speed for real world usage (especially for static images) and is hence recent go-to method.

A further analysis of the recent method (accurate path) for text extraction in general use cases (not specifically targeted to our application) by running it on multiple datasets including the ICDAR 2003 (IC03) [19], IIIT5K [20] and Street View Text (SVT) [1] as shown in the 8

comparison Table 2. The SVT dataset consists of 647 images from the Google Street view (cropped), IC03, the IIIT 5K word dataset, which includes both full scene images251 and cropped images of words (860), and 5000 clipped word images from scene texts. It is obtained from Google image search using query words such as signs, billboards, name boards of houses, house numbers, movie posters, used to collect images. Full scene images are used for our tests on the IC03, as they are more relevant to the current application.

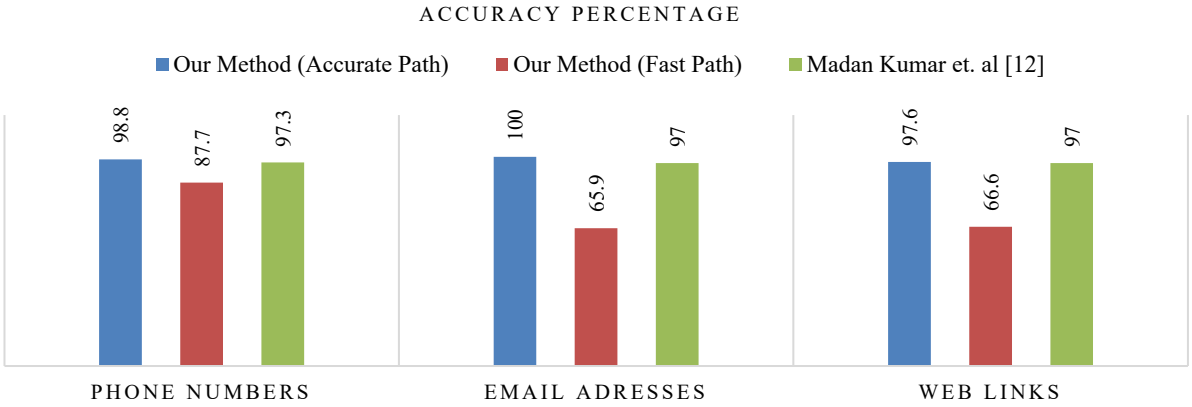


Figure 7. Comparison of text extraction and classification accuracy for various data types

It is found that the applied method gives promising results in the above-mentioned standard, general purpose text on image datasets, and sits in the league of top-of-the-line algorithms that have been specifically designed for them. The results of the testing are shown in Figure 8. Based on the experimental test results of different methods, it is found the accurate path to be the most suited method for practical usage of dynamic images providing highest accuracy for all of the relevant categories, top of the line overall accuracy, and extremely fast execution time of 400 ms on average.

Table 2. Text recognition accuracy comparison of different papers on multiple datasets

Reference (Relevant Papers)	SVT [16]	IC03 [17]	IIIT5k [18]
B. Shi, et al. [7]	94.4	89.4	80.8
B. Shi, et al. [8]	93.8	90.1	81.9
Z. Cheng, et al. [9]	82.8	91.5	87.0
Lee, et al. [10]	80.7	88.7	78.4
M. Jaderberg, et al. [11]	71.7	89.6	-
Proposed Method (Accurate Path)	88.7	89.2	82.5

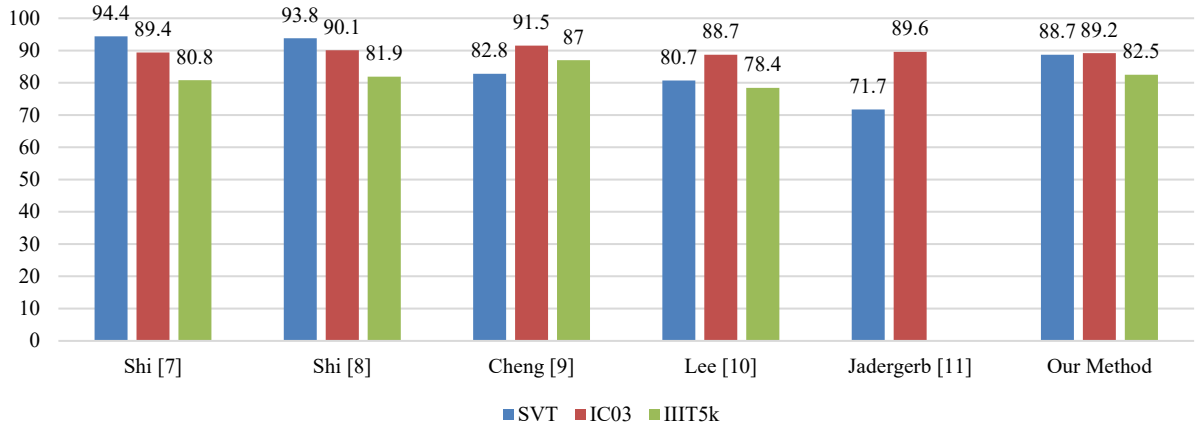


Figure 8. Comparison of text extraction for multiple datasets

5. CONCLUSION

In this study, it is shown in this implementation to dynamically visualize an image to enable direct interaction with the phone numbers, email addresses, and web links on it. This has been achieved by using state-of-the-art technologies including the Apple Developer's Vision framework for text extraction and localization, NSDataDetector API for data classification, and platform specific actions implemented using deep linking through the open API.

It is implemented and compared two different paths for text extraction: accurate path and fast path, and have found the accurate path to be the most suited method for practical usage where accuracy is the most important factor. The proposed method (accurate path) shows promising results and comes out at the top by leading the accuracy in all the specific data type cases with the average accuracy of 98.8% on the Stanford Mobile Visual Search Dataset for Business Cards along with a very fast execution time of 400 ms per image.

The proposed method also gives top of the line results for general text extraction on datasets including the SVT and IC03. With digital transformation of information and the increasing usage of images, the technology described above is a significant utility for added convenience, safety, and efficiency.

REFERENCES

[1] V. Jain, Y. Jain, H. Dhingra, D. Saini, M.C. Taplamacioglu, M. Saka, "A Systematic Literature Review on QR Code Detection and Pre-Processing", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 46, Vol. 13, No. 1, pp. 111-119, March 2021.

[2] N. Atashafrazeh, A. Farzan, "A Review of Using Machine Learning Algorithms for Image Retrieval Words", International Journal on Technical and Physical Problems of Engineering (IJTPE), Issue 20, Vol. 6, No. 3, pp. 139-144, September 2014.

[3] "Vision Framework", Apple Developer Documentation, 3 June 2021, https://developer. apple.com/documentation/vision.

[4] "NSDataDetector", Apple Developer Documentation. https://developer.apple.com/documentation/foundation/nsdatadetector, 3 June 2021.

[5] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, J. Lian, "EAST: An Efficient and Accurate Scene Text Detector", 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2642-2651, 2017.

[6] Z. Zhang, C. Zhang, W. Shen, C. Yao, W. Liu, X. Bai, "Multi-Oriented Text Detection with Fully Convolutional Networks", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4159-4167, Las Vegas, NV, USA, June 2016.

[7] B. Shi, X. Bai, C. Yao, "An End-to-End Trainable Neural Net-work for Image-Based Sequence Recognition and Its Application to Scene Text Recognition", IEEE Transactions on Pattern Analysis Machine Intelligence, Vol. 39, No. 11, pp. 2298-2304, 2017.

[8] B. Shi, X. Wang, P. Lyu, C. Yao, X. Bai, "Robust Scene Text Recognition with Automatic Rectification", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 4168- 4176, June 2016.

[9] Z. Cheng, Y. Xu, F. Bai, Y. Niu, S. Pu, S. Zhou, "AON: Towards Arbitrarily-Oriented Text Recognition", IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5571-5579, Salt Lake City, UT, USA, June 2018.

[10] C. Lee, S. Osindero, "Recursive Recurrent Nets with Attention Modeling for OCR in the Wild", 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2231-2239, Las Vegas, NV, USA, June 2016.

[11] M. Jaderberg, K. Simonyan, A. Vedaldi, A. Zisserman, "Deep Structured Output Learning for Unconstrained Text Recognition", Computer Science, December 2014.

[12] C. Madan Kumar, M. Brindha, "Text Extraction from Business Cards and Classification of Extracted Text into Predefined Classes", International Journal of Computational Intelligence IoT, Vol. 2, No. 3, 2019.

[13] A. Uysal, S. Gunal, "The Impact of Preprocessing on Text Classi- Fication", Information Processing Management, Vol. 50, No. 1, pp. 104-112, January 2014.

[14] "Core Graphics Framework", Apple Developer Documentation.https://developer.apple.com/documentation/coregraphics, 03 June 2021.

[15] "Open URL API, UIKit", Apple Developer, https://developer.apple.com/documentation/uikit/uiapplication/1648685-open, 3 June 2021.

[16] V. Chandrasekhar, D. Chen, S. Tsai, N. Cheung, H. Chen, G. Takacs, Y. Reznik, R. Vedantham, R. Grzesczcuk, J. Bach, B. Girod, "The Stan- Ford Mobile Visual Search Dataset: Business Cards", The First ACM Multimedia Systems Conference (MMSys), San Jose, CA, USA, 23-25 February 2011.

[17] Michelle Ehrhardt, "Apple M1 Chip: Specs and Performance", Tom's Hardware, www.tomshardware.com/news/Apple-M1-Chip-Everything-We-Know, 3 June 2021.

[18] K. Wang, B. Babenko, S. Belongie, "End-to-End Scene Text Recognition", International Conference on Computer Vision, pp. 1457-1464, Barcelona, Spain, November 2011.

[19] S.M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, R. Young, "ICDAR 2003 Robust Reading Competitions", Seventh International Conference on Document Analysis and Recognition, pp. 682-687, Edinburgh, Scotland, August 2003.

[20] A. Mishra, K. Alahari, C.V. Jawahar, "Scene Text Recognition using Higher Order Language Priors", BMVC 2012, https://cvit.iiit.ac.in/research/projects/cvit-projects/the-iiit-5k-word- dataset.

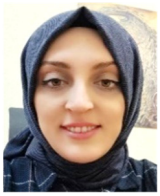
BIOGRAPHIES

Vanita Jain was born in New Delhi, India in 1966. She received her B.E. degree in Electrical Engineering. in 1984, M.Tech degree in Controls in 1990 and Ph.D. Degree from V.J.T.I., Mumbai, India. She is the first woman to have completed Ph.D. in Technology from Mumbai University, India. She has more than 30 years of teaching experience, and taught students at both UG and PG level. Currently, she is working as Professor and Head of Information Technology Department at Bharati Vidyapeeth's College of Engineering, New Delhi, India since 2010. Her field of interest includes soft computing, control, optimization techniques and system engineering and is actively involved in the research in these areas.

Yugantar Jain was born in New Delhi, India in 2000. He is currently working as a software engineer in Eimy GmbH. He received his B.Tech. degree in the field of Information Technology from Bharati Vidyapeeth's College of Engineering, New Delhi, India. His areas of research are QR codes, dynamic images, text extraction, text classification and data detection.

Hardik Dhingra was born in Haryana, India in 1999. He graduated Bachelor degree in Technology in Information Technology from Bharati Vidyapeeth's College of Engineering, New Delhi, India in 2021. His research interests and subjects are QR codes, dynamic images, text extraction, text classification and data detection.

Achin Jain was born in Delhi, India, in 1984. He received the M.Tech. degree in Information Security in 2013 and B.Tech. degree in Information Technology from GGSIPU, Delhi, India in 2007. He is currently a Ph.D. student of University School of Information, Communication and Technology, GGSIPU, and works at Bharati Vidyapeeth's College of Engineering, New Delhi, India as Assistant Professor since 2014. His research interests are sentiment classification and machine learning using NLP techniques. He has published more than 15 articles in peer reviewed journals and presented work in more than 7 international conferences.



Merve Demirci was born in Trabzon, Turkey in 1992. She graduated from Departmen of Electrical and Electronics Engineering, Ataturk University, Erzurum, Turkey in 2014 and working as a Research Assistant at Gazi University, Ankara, Turkey. She received the degrees of M.Sc. degree and currently a student of Ph.D. degree in Electrical and Electronics Engineering Department, Gazi University. Her main research interests and subjects are power systems analysis, artificial intelligence and power transformer fault analysis.

M. Cengiz Taplamacioglu was born in Ankara, Turkey in 1962. He graduated from Department of Electrical and Electronics Engineering, Gazi University, Ankara, Turkey. He received the M.Sc. degrees in Industrial Engineering from Gazi University and also in Electrical and Electronics Engineering from Middle East Technical University, Ankara, Turkey. He received his Ph.D. degree in Electrical, Electronics and System Engineering from University of Wales, Cardiff, UK. He is a Professor of the Electrical and Electronics Engineering since 2000. His research interests and subjects are high voltage engineering, corona discharge and modelling, electrical field computation, measurement and modelling techniques, optical HV measurement techniques, power systems control and protection, lighting techniques, renewable energy systems and smart grid applications.