International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 9 Issue 9 ǁ 2021 ǁ PP. 20-27

Optical Character Recognition using Natural Images based on ResNet50 Pre-trained model with Data Augmentation

Hemant Bhiwapurkar
M. Tech. Scholar, All Saints' College of Technology, Bhopal, India

Zuber Farooqui
Professor, Dept. of CSE, All Saints' College of Technology, Bhopal, India

Abstract: *In the past, text detection and recognition have been a good issue. However, detecting and identifying text in natural images and natural scenes becomes a significantly more difficult task due to the distortion in geometry and variation in light. Deep learning algorithms have recently reached state-of-the-art object identification and recognition performance. On the other hand, deep learning needed a vast data collection and a lot of processing power to build a model from scratch. Therefore, in this research work, a deep learning technique is used, based on the knowledge transfer of pre-trained Convolutional neural networks(CNNs) to recognize the text in the natural images.*

*In this work, we have used a pre-trained ResNet50 model with data augmentation. To fine-tuned the ResNet50 model top layer is removed, and a new layer has been included in the network then it is trained (fine-tuned) on the optical character images. A CHAR74K dataset, which is a benchmark for natural images, is used to evaluate the proposed method**.** Data augmentation methods such as rotation and flipping are used to improve the performance of the proposed model. The proposed model has achieved an accuracy of 87.56 % and an F1-score of 88%. The proposed model compares with state of the art method and achieves better performance.*

Keywords: Convolutional Neural Networks (CNN), Deep Learning, Optical Character Recognition (OCR), Maximally Stable External Region (MSER), pre-trained network ResNet50.

Date of Submission: 09-09-2021 Date of acceptance: 24-09-2021

I. Introduction

In vision-based applications, text is very essential. Text detection is the process of identifying text in a set of images. Text reading in natural images is critical in most modern computer vision applications, such as visual assistance and image retrieval. Because writing in images generally conveys significant information, this is the case. As a result, text recognition and identification has gotten a lot of attention in recent years.

Text recognition and identification in natural pictures has been a significant and difficult topic in recent years. It also has a wide range of uses. Unlike character identification in scanned documents, identifying text in natural and unconstrained pictures is a difficult task exacerbated by typefaces, lighting conditions, textures, and a wide range of backdrop changes. Unfortunately, text characters in natural pictures can be of any gray-scale hue (even white), vary in length, have low resolution, and appear in complicated backdrops..

An image of nature Because image text often contains useful information, text reading is crucial for many sophisticated applications such as image and video collecting, situation interpretation, and visual assistance. It is critical to read material. As a result, the identification and detection of text in scene photographs has become increasingly important in modern society. Despite significant research in recent years, text identification in unconstrained settings remains challenging due to a variety of variables, including high variability in character font, width, colour, and orientation, as well as complex context and non-uniform illumination..

Previous works in that area were used for text detection based on sliding windows and the study of the related components. Through rotating a multi-scaled grading panel, sliding window-based methods identify text areas. Although this exhaustive search achieves high recall levels, it is computer-inefficient. Methods focused on connected components derive characteristics through the analysis and the technique for grouping and refining of connected components. False alarm can therefore be removed and non-text elements eliminated. Stroke width transform (SWT)[1] and MESR[2] are two representative methods that achieve state-of - the-art output on a CHAR74 K dataset with MSER-based methods[3]. The MSER algorithms, however, extract massive non-text repetitive parts which are restricted by the false deletion and refining rules. These techniques can not also detect noise or distorted characters in the background. More recently, many deep learning-based approaches were developed for text scene detection due to profound model functionality representations. These models calculate high-level deep features from image patches or proposals for text / non-text classification based on convolutional Neural Networks (CNNs). These methods are also limited by the discriminative capacity of the local proposals and the CNN classifiers. In this project we suggest a flexible solution that incorporates the advantages of MSER and CNN characteristics. We may summarize our contributions in three points. First, an enhanced saliency-MSER is suggested, which is an extension of a well-known MSER algorithm by adding saliency-detection methods on three channels of the object as a characteristic candidate extractor. The pipeline with deep CNN message sorting is the second. We train a powerful convolutional neural network that includes pixel and character level data in the classification stage.

II. Literature Survey

Character recognition is not a new problem; it may be traced back to structures that predate computer invention. The original OCR systems weren't computers, but mechanical machines that could recognise characters, although they were sluggish and inaccurate. M. M. Sheppard built a GISMO reading and robot in 1951, which is considered the first achievement in modern OCR [11]. GISMO will read musical notes and text one by one on a printed page. However, there are only 23 characters that can be recognised.. The computer can also copy the typewritten page from the machine. J. A computer that could read uppercase typed English characters one per minute was developed by Rainbow, in 1954. The early OCR systems were criticized due to errors and slow recognition speed. Hence, not much research efforts were put on the topic during 60’s and 70’s. The only developments were done on government agencies and large corporations like banks, newspapers and airlines etc. Because of the complexities associated with recognition, it was felt that three should be standardized OCR fonts for easing the task of recognition for OCR. Hence, OCRA and OCRB were developed by ANSI and EMCA in 1970, that provided comparatively acceptable recognition rates[12]. During the past thirty years, substantial research has been done on OCR. This has lead to the emergence of document image analysis (DIA), multi-lingual, handwritten and omni-font OCRs [12]. Despite all of these improvements, the machine's ability to accurately comprehend text is still considerably inferior to that of a person. As a result, current OCR research is focused on increasing the accuracy and speed of OCR for a variety of printed and written texts in unconstrained situations. For complicated languages such as Urdu or Sindhi, there hasn't been any open source or commercial software accessible.

In order to extract text and symbols with great confidence, the pre- processing of the image is what matters. Pre-processing within the engines is most often limited, if too much pre-processing is done on the picture it may not be as the developer expected, or it might be a strategic move to tell developers to do their own pre-processing before using the OCRengine.

The following tasks should be performed by a general OCR system including pre-processing[13]:
1. Text digitization
2. Noise clarification
3. Text Block for Identification
4. Line and Word Detection
5. Character segmentation
6. Feature extraction
7. Character Recognition
8. Error Correction

This is the recommended order for evaluating an image with good results, however the conclusion is dependent on the situation or purpose of the image. It is common to begin preprocessing an image before going on to the next step[14].The first steps to transform the image into a gray image. Nevertheless, another method can be used to find specific colored areas in an image by using color binarization. It works by setting the color hue limit above and below. For all pixels in a picture between the threshold, 1 (white) is 0 (black) otherwise[15], 0 is converted. I've done the work. Kastelanetal.[16] have been perfectly aligned in one of the thesis goals to evaluate the extracted text. The Tesseract OCR engine is said to produce good results. They plan to improve the system in the future to better recognize text regions, as issues emerge when the engine interprets the symbols. A lower score is assigned.

'There is no need for OCR systems that handle special symbols to encourage further growth,' according to J. Liang et al.[17]. The database's efficiency will be hampered if there are too many symbols saved. The challenge in implementing such a function is making the system a universal solution. The most popular method for identifying symbols is to scan documents for mathematical symbols. Tesseract funded the development of a mathematical language that understands symbols. This is a clever technique to capture both natural and mathematical symbols using the OCR engine.

OCR may be used to recognize symbols in road surface mapping applications[18]. According to the study, the procedure for identifying road signs is the same as the facial recognition process (a device that recognises faces). Tesseract was used as a selection motor once again by the researchers. The experiment using road symbols yielded 80 percent properly classified symbols and 2.5 percent inaccurately labelled symbols.

The study, found Tesseracts to be ideal with high resolution and sharp images, yielded good results despite the poor quality of the camera and the vehicle was moving during the trial. The more complex symbols with the same contour and different interior were the most difficult symbols to be recognized. A separate study was required in order to distinguish symbols with the same layout with different insides. Even signs of distinct outlines could be identified and analysed.

Tesseract began as a PhD research project at HP Laboratories in Bristol. It proved popular, and HP produced it between 1984 and 1994. In 2005, Tesseract was made available as open source software[19]. Instead, the classifier is used to identify broken characters and replace them completely. Damaged data isn't recognised by the engine since it hasn't been programmed to recognise it. This might result in a significantly smaller and more efficient learning sample repository. It started off with 20 samples of 94 distinct characters, eight different typefaces with 4 distinct features, and a maximum sample size of 60160 samples (normal, Bold, Italic, Bold Italic).

During the OCR engine evaluation, Tesseract was referenced several times. Signs [18] are also utilised for analysis unless the tests are correct [20]. The engine is promising, but it isn't robust enough to draw particular conclusions. Tesseract [21] supports UTF 8 and recognises over 100 languages from the container, with additional languages on the way. Because the motor may be trained, it is feasible to learn and understand a new language or font that is not typically promoted. Tesseract is regarded as the most accurate open source motor on the market. Leptonicautilises the image care library [22] for the fundamental collection.The software is created and built in C or C++, but other programs can be used in other languages. The Apache license 2.0[23] is valid. A third-party wrapper: JavaCPP along with its various presents [24] is used to work with Tesseract inJava.

Tan Chang Wei at el [25], used deep neural network for the learning and execution of OCR using Inception V3. The V3 network Inception consists of 53,342 noisy pictures from receives and newspapers. Fabio De Sousa Ribeiroat al. [26] a dual deep, neural network-based system for automatically identifying dates of usage on the food package photos is proposed for an end-to-end architecture. The system includes: a global neural convolution (CNN) network to assess the value of food packets (blurry / clear / missing use by date data); and a completely convolutional network at the local level (FCN) for ROI position to use bydate.

Choudhary Savita at the. 27] Proposes a text area detection method with MSRs and a self-trained neural network to recognizes the message. MSR theory. MSR theory. The photo is prefabricated, MSER and the canny edge are used to find the smaller areas that can more likely contain text. The text is extracted as a single character by simple algorithms in the binary image and the method for recognition is passed through which dim characters are specificallydesigned.

Document recognition is where OCR systems come in very handy because it's impossible to regulate the production process. This can occur if the receiver is isolated from an electronic version and does not supervise the manufacturing process, or if the receiver is given outdated content that cannot be digitally processed at the time of creation. Future OCR-systems must be connivant in order to read the imprinted text. Another major area for OCR is the identification of hand-produced documents. In the area of postal applications, OCR will focus on reading addresses in email from those who do not have access to computer software. It's no longer rare for companies, etc. to mark mails with barcodes, with access to computer technology. The relative importance of handwritten text recognition is therefore expected toincrease.

III. Convolutional Neural Network (CNN)

Convolutional neural network (ConvNets or CNNs) is one of the most important categories for the recognition of objects. Detection of objects, reconnaissance faces etc. are some of the areas in which CNNs are used.

Input images are taken, processed and classified in certain categories of CNN image classifications (eg., Dog, Cat, Tiger, Lion). Computers interpret the input image as a pixel array and the resolution depends on it. The image resolution shows hx wxd(h= large, w= wide, d= size). Based on

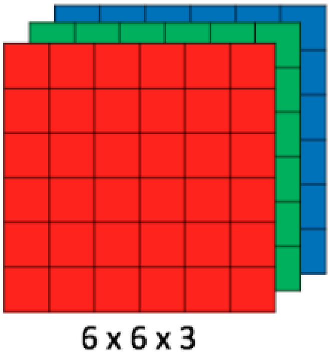

Figure 1: Matrix of grayscale image

the image resolution. The RGB matrix image (3 refers to values in RGB) is 6x 6x 3 and the grayscale matrix image is 4x 4x 1.

That input image must pass a series of convolution layers of CNN models (kernels), pooling, fully-connected (FC) layers to theoretically deeply understand, and use Softmax to classify an object with probabilistic values from zero to 1. The following figure represents a complete CNN flow for input picture processing and values classification.

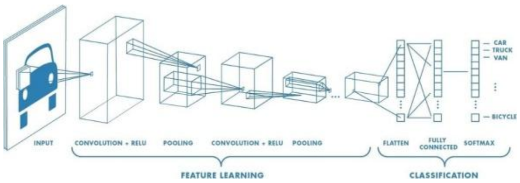

Figure 2: CNN architecture

4.ResNet Architecture

ResNet is a pre-trained network that is trained on the ImageNet database by considering millions of images of large varieties, containing about 1.4 million images and 1000 classes. The name ResNet50 [8] is a short form of Residual Neural Network. It is a 177-layer convolutional neural network with residual blocks

Architecture of ResNet50 is shown in the figure 3.12 and residual block is represented in figure 3.13. The main idea of ResNet is based upon “identity shortcut connection” that involves skipping one or more layers from the network as shown in the figure 4.3.

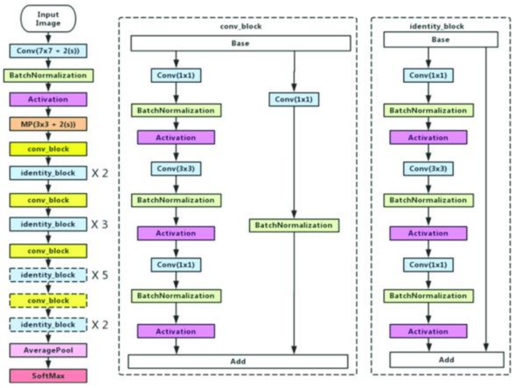

Figure 3: General Architecture of the ResNet50.

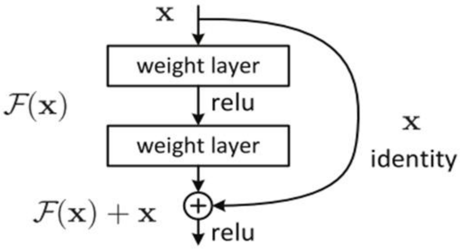
x
weight layer
$\mathcal{F}(\mathbf{x})$
relu
weight layer
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 4: Residual Block of the ResNet50.

Over the time, new convolutional neural network (CNN) models have been coming up, feature extraction is made possible using CNN models and many computer vision related tasks too. Even though, the standard CNN models that are widely used these days have a really high computational cost and require high capacity for storage & memory, which is impractical and very expensive for commercial applications such as real-time on-road object detection on embedded boards or mobile platforms.

IV. Proposed Methodology

In the proposed method transfer learning is used, in which network architecture and network`s weighs of retrained network ResNet50 is directly used for fundus image classification. To improve the performance of the proposed approach the data augmentation method has been applied. Then network is fine-tuned according our application on the fundus images. Following section gives the details descriptions of the proposed method:

5.1 Data augmentations

We uses affine transformation based data augmentation technique to improve the performance of network by overcome the problem of network overfitting [28]. Data augmentation method is only applyied on the training set and for validation set and testing set real world data have used. For trainimng set augmentation first we rotated the images by 90^0 , 180^0 and 270^0 , then flip opration is applied in which pixels are flip with level and vertical direction and at last translation is applied

To fine-tune the network following algorithm 1 is applied:

Algorithm 1: Fine tuning pre-trained ResNet50 network

Input: Train set and validation set

Output: Trained model for natural image classification

Begin

Step 1. Remove fully connected layer, softmax layer and output layer of the network

Step 2. Add new fully connected layer with 62 neurons (for multiclass classification), softmax layer and output layer.

Step 3. Set high learning rate for newly added layers and very low learning rate for the remening.

Step 4. Re-trained the nework on the new application dataset (natural image char74k dataset).

5.3 Optical Character Recognition and classification algorithm

We have use concept of transfer learning to detection and classification the optical character recognition from the natural images. In the proposed approach pre-trained ResNet50 model is fine-tuned by using train set of the natural images char74k dataset. Test set images are given as the input to the trained model, which produces classify images into the 62 classes. Proposed approach algorithm is given as follows:

Algorithm 2: Classification of Natural images

Input: Query and database images

Output: Retrieved Images

Begin

Step 1. Read all images from database.

Step 2. Divide dataset into train set, test set and validation set.

Step 3. Apply data augmentation approach on the train set.

Step 4. Fine-tuned pre-trained ResNet50 model by applying ***Algorithm 1*** (pass train set and validation set in the input).

Step 5. Now read test images.

Step 6. Test images are given as input to the trained model and get label for each input image as output.

End

VI. Implementation and Result Analysis

6.1 Dataset

Character recognition is a classic pattern recognition problem that scientists have been working on since computer vision's early days. With today's cameras omnipresence, automatic character recognition applications are wider than ever before.

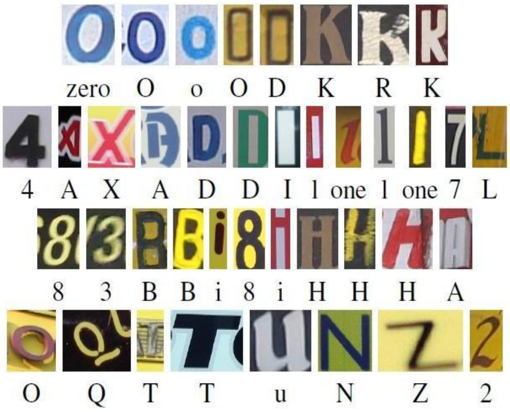

Figure 5: Sample image of Char74K dataset

This is largely considered a solved problem in restricted circumstances of Latin script, such as images of scanned documents with common character fonts and uniform context. Photos captured with common cameras and hand held phones, however, still present a formidable challenge for the identification of characters. In this data set, the challenging aspects of this issue are evident. Char74K consist 7,712 optical character images of 62 classes. In which 10 classes of numbers from 0 to 9 and 52 classes of 26 lower case 26 upper case characters. Sample image of Char74K dataset is shown in the figure 5.1.

We have divided Char74K dataset randomly into a training set, validation set and test set in the ratio of 7:20:10. Number of images presents in the each set is given in the Table 1 and image distribution in each set corresponding to each class is shown in the figure 4.2. Training set is used to train the ResNet50 model, validation set is used for selecting based parameters of the network and test set is used for performance analysis of the proposed model.

Table 1: Number images in the training, test and validation set.

Set	**Total number of images**
Training Set	5403
Test set	1543
Validation	766

6.2 Result Analysis

In this section, we present the results obtained from the experiment over the database of Char74 dataset. Table 2 shows the performances of the model on the training, validation and test set. As observed from the table network performance on the validation set and test are the approximately same, therefore we can say our model is stable.

Table 2: Network performance of the Char74K dataset.

Set	**Accuracy (in %)**
Training Set	93.24
Test set	87.56
Validation	88.65

Table 7: Proposed Model performance comparison with Existing method

Method	**Accuracy (in %)**	**Dataset**	**No. of classes**
CNN [30]	83.00	ICDAR2003	49
KNN [31]	35.47	Chars74k	62
Linear Classifier [31]	30.15	Chars74k	62
LeNet [31]	45.36	Chars74k	62
AlexNet [31]	63.38	Chars74k	62
AlexNet [32]	77.77	Chars74k	62
ResNet50 +Data Augumentation(Proposed method)	87.56	Chars74k	62

VII. Conclusion

Text detection and recognition has previously been a well-studied topic. However, due of the distortion in geometry and variation in light, detecting and identifying text in natural pictures and text from natural settings becomes a considerably more difficult task. This research work presented a transfer learning and data augmentation-based approach for optical character recognition method. It is capable of differentiating the optical character from the natural scene image. In study, we have used the architecture and weights of pre-trained ResNet50 model. We have first fine-tuned the proposed model on the optical character Char74K dataset. To fine-tuned the ResNet50 model top layer is removed and new layer have been included in the network then it is trained (fine-tuned) on the optical character images. Our proposed system exhibits good performance over CHAR74K character dataset and it has achieved an accuracy of 87.56 % and F1- score of 88%.

Future Scope

The minimal amount of training data is one of the model's main drawbacks. We experimented on a variety of pictures and found that the classifications were sometimes inaccurate. When this happened, we looked into the input picture network more and discovered that the most dominating color(s) in the image have a significant impact on categorization. Over the years, character recognition methods have improved from quite primitive systems, which are only suitable for the reading of stylish printed numerals to more complex and advanced techniques for recognizing a wide range of typeset fonts and man-made characters. As computer technology advances and computer limitations decrease, new techniques of character recognition are predicted. On grey level images, for example, it may be possible to perform character recognition directly. However, by integrating methodologies and allowing more use of contexts, the greatest potential is the manipulation of existing methods. Integrating segmentation and contextual analysis can improve character recognition.

References

[1]. Epshtein, Boris, EyalOfek, and Yonatan Wexler. "Detecting text in natural scenes with stroke width transform." In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 2963-2970. IEEE,2010.

[2]. Gou, Chao, Kunfeng Wang, Zhongdong Yu, and HaitaoXie. "License plate recognition using MSER and HOG based on ELM." In Proceedings of 2014 IEEE International Conference on Service Operations and Logistics, and Informatics, pp. 217-221. IEEE,2014.

[3]. De Campos, T., R. B. Bodla, and M. Varma. "The chars74k dataset."(2009).

[4]. Neubeck, Alexander, and Luc Van Gool. "Efficient non-maximum suppression." In 18th International Conference on Pattern Recognition (ICPR'06), vol. 3, pp. 850-855. IEEE, 2006.

[5]. Richard Szeliski. Computer Vision - Algorithms and Applications. Texts in ComputerScience. Springer, 2011.

[6]. Ding Liu. Connecting low-level image processing and high-level vision via deep learning. In Proceedings of the Twenty-Seventh International Joint Conference on ArtificialIntelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden., pages 5775-5776,2018.

[7]. Maya R. Gupta, Nathaniel P. Jacobson, and Eric K. Garcia. OCR binarization and image pre-processing for searching historical documents. Pattern Recognition, 40(2):389-397,2007.

[8]. Derek Bradley and Gerhard Roth. Adaptive thresholding using the integral image. Journal of graphics tools, 12(2):13{21, 2007.

[9]. Abdellatif Abdelfattah. Image classification using deep neural networks - a beginner friendly approach using tensorflow.https://medium.com/@tifa2up/imageclassification-using-deep-neural- networks-a-beginner-friendlyapproach-using-tensorflow-94b0a090ccd4.[2019-02-26].

[10]. Lars Hulstaert. https://www.datacamp.com/community/tutorials/ objectdetection-guide, 4 2018.

[11]. D. C. Ciresan, U. Meier, J. Masci, L. M. Gambardella, and J. Schmidhuber. High performance neural networks for visual object classification. Technical Report IDSIA-01-11, DalleMolle Institute for Artificial Intelligence, 2011.

[12]. A. Coates, B. Carpenter, C. Case, S. Satheesh,B. Suresh, T. Wang, D. J. Wu, and A. Y. Ng. Text detection and character recognition in scene images withunsupervised feature learning. In ICDAR, 2011.

[13]. Divakar Yadav, Sonia S´anchez-Cuadrado, and Jorge Morato. Optical character recognition for hindi language using a neural-network approach. JIPS, 9(1):117{140, 2013.

[14]. Aamir Khan, Devesh Pratap Singh, Pushpraj Singh Tanwar, and Amit Rajput. Vehicle license plate number recognition and segmentation system. International Journal on Emerging Technologies, 2(2):75{79, 2011.

[15]. Muhammad Tahir Qadri and Muhammad Asif. Automatic number plate recognition system for vehicle identification using optical characterrecognition. In Education Technology and Computer, 2009. ICETC'09. International Conference on, pages 335-338. IEEE, 2009.

[16]. Ivan Kastelan, Sandra Kukolj, VukotaPekovic, Vladimir Marinkovic, and Zoran Marceta. Extraction of text on tv screen using optical character recognition. InIntelligent Systems and Informatics (SISY), 2012 IEEE 10th Jubilee International Symposium on, pages 153{156. IEEE, 2012.

[17]. Jisheng Liang, Ihsin T Phillips, Vikram Chalana, and Robert Haralick. A methodology for special symbol recognitions. In Pattern Recognition, 2000. Proceedings. 15thInternational Conference on, volume 4, pages 11-14. IEEE, 2000.

[18]. Markus Schreiber, Fabian Poggenhans, and Christoph Stiller. Detecting symbols onroad surface for mapping and localization using ocr. In Intelligent Transportation Systems (ITSC), 2014 IEEE 17th International Conference on, pages 597-602. IEEE, 2014.

[19]. Ray Smith. An overview of the tesseract ocr engine. In icdar, pages 629-633. IEEE, 2007.

[20]. Ocr software comparison tesseract vs abbyyfinereader, 2015. http://lib.psnc.pl/ dlibra/docmetadata?id=358&from=publication&showContent=true.Accessed: 2016-02-27.

[21]. Tesseract optical character recognition engine. https://github.com/tesseract-ocr. Accessed: 2016-02-05.

[22]. Leptonica, useful for image processing and image analysis applications. http://www.leptonica.com/.Accessed: 2016-04-06.

[23]. Apache license 2.0. http://www.apache.org/licenses/LICENSE-2.0. Accessed: 2016-02-26.

[24]. Javacpp- presets module contains java configuration and interface classes for widelyused c/c++ libraries. https://github.com/bytedeco/javacpp-presets. Accessed: 2016-02-26.

[25]. Wei, Tan Chiang, U. U. Sheikh, and Ab Al-Hadi Ab Rahman. "Improved optical character recognition with deep neural network." In 2018 IEEE 14thInternational Colloquium on Signal Processing & Its Applications (CSPA), pp. 245-249. IEEE, 2018.

[26]. Ribeiro, Fabio De Sousa, Liyun Gong, Francesco Calivá, Mark Swainson, KjartanGudmundsson, Miao Yu, Georgios Leontidis, Xujiong Ye, and Stefanos Kollias. "An end-to-end deep neural architecture for optical character verification and recognition in retail food packaging." In 2018 25th IEEE International Conference on Image Processing (ICIP), pp. 2376-2380. IEEE, 2018.

[27]. Choudhary, Savita, Nikhil Kumar Singh, and Sanjay Chichadwani. "Text Detection and Recognition from Scene Images using MSER and CNN." In 2018 Second International Conference on Advances in Electronics, Computers and Communications (ICAECC), pp. 1-4. IEEE,2018.

[28]. T. E. de Campos, B. R. Babu and M. Varma.Character recognition in natural images. In Proceedings of the International Conference on Computer Vision Theory and Applications (VISAPP), Lisbon, Portugal, February2009.

[29]. Kosub, Sven; "A note on the triangle inequality for the Jaccard distance" arXiv:1612.02696

[30]. Ghani, Rana Fareed. "Robust character recognition for optical and natural images using deep learning." In 2019 IEEE Student Conference on Research and Development (SCOReD), pp. 152-156. IEEE, 2019.

[31]. Sundaresan, Vishnu, and Jasper Lin. "Recognizing handwritten digits and characters." (1998).

[32]. Soomro, Moazam, Muhammad Ali Farooq, and Rana Hammad Raza. "Performance evaluation of advanced deep learning architectures for offline handwritten character recognition." In 2017 International Conference on Frontiers of Information Technology (FIT), pp. 362-367. IEEE, 2017.