Urdu-Text: A Dataset and Benchmark for Urdu Text Detection and Recognition in Natural Scenes

Asghar Ali
School of Engineering and Information Technology,
University of New South Wales,
Canberra, Australia
Email: a.chandio@student.adfa.edu.au

Mark Pickering
School of Engineering and Information Technology,
University of New South Wales,
Canberra, Australia
Email: m.pickering@unsw.edu.au

***Abstract*—Multi-lingual text in natural scene images conveys useful information and is a fundamental tool for tourists to interact with their environment. Multi-lingual text detection and recognition in natural scenes, therefore, has become a challenging problem for researchers in the last few years. Recently, a large-scale multi-lingual dataset for scene text detection and script identification is published by the ICDAR which, contains scene images with text in six different scripts including Arabic. This paper presents a novel dataset and benchmark for Urdu text in natural scenes. Currently, no dataset for Urdu text in natural scenes is publicly available. Urdu is a type of cursive language, which is derived from Arabic script and uses many similar alphabet characters. Therefore, the proposed dataset could be helpful for multi-lingual text detection, recognition and script identification. The aim of this dataset is to help the research community for algorithm development and evaluation of Urdu text in natural scenes. The Urdu-Text dataset contains 1400 complete scene images and 8200-segmented words. The images in the dataset contain a broad variety of text instances in multi-orientations with small and large font sizes. The dataset contains ground truths in the form of bounding boxes at the word level, the script of the text and the text-transcription. The performance of three deep neural networks is evaluated to measure the robustness of the Urdu-Text dataset.**

***Index Terms*—Urdu Scene Text, Urdu Natural Image Text Dataset, Urdu Text Recognition, Segmentation-based Words**

I. Introduction

Reading multi-lingual text in natural scene images remains a complex and challenging task in computer vision and document analysis research. The ability to detect and recognize multi-lingual text in natural scene images could be very helpful in various real-world applications such as translation for foreign tourists, script identification, geo-localization, product label recognition and understanding ATM machine instructions. In recent years, many well-known public datasets such as ICDAR 03, 05, 11, 13, 15 [1], MSRA-TD500 [2], Street View Text [3], COCO-Text [4], Total-Text [5] and others have played an important role in initiating natural scene related research. For Latin text, the problem has been well studied and state-of-the-art techniques have been proposed [6]–[9]. However, much less concentration has been given to multi-lingual text detection and recognition in natural scenes. In particularly, the work done for text detection in cursive languages such as Arabic and its derived languages is at infancy stage. The development and evaluation of state-of-the-art methods typically require a dataset with hundreds of images containing text instances. Recently, the International Conference on Document Analysis and Recognition (ICDAR) has initiated a competition on multi-lingual text detection and script identification in scene images where, a dataset with complete scene images containing text in nine different languages represented with six scripts is published [10]. This dataset contains a total of 7200 training samples, which are equally divided into the same number of images per script, i.e. 800 training samples for each of the following languages: Arabic, English, Chinese, Bangla, Korean, German, Italian, Japanese and French. So far, this is the largest multi-lingual natural scene text dataset. The number of training samples for each script are very small and are insufficient to train very deep models. The Urdu-Text dataset proposed in this paper would be an extension in the resources of ICDAR MLT dataset [10].

Urdu is a type of cursive text where the letters are connected and is written from right to left. It is derived from Arabic script, therefore, most of the alphabets of Urdu script are the same as for Arabic script. Both languages can be categorized based on writing style. Urdu script is usually written in Nastaliq style, whereas Arabic script is written in Naskh style. Both are cursive languages and have similar challenges and difficulties in natural scene text detection and recognition. Generally, text in Arabic and Urdu natural scene images is diverse ranging from very small to very large font sizes, machine printed and handwritten. Figure 1 shows some examples from the Urdu-Text dataset embedded with diverse text sizes, machine printed and handwritten text, annotation details and associated challenges to detect and recognize the text correctly.

The primary contribution of this research is to develop an Urdu-Text dataset which, will help the research community to evaluate existing methods and develop new state-of-the-art techniques for multi-lingual natural-scene text detection and recognition. The Urdu-Text contains 1400 complete scene images with Urdu, English and some Sindhi as well as Arabic text instances. The text instances in the images have varieties from machine-printed text on shop names, advertisement banners or street names to handwritten text written on the walls and hoardings. The dataset is split into 1000 training samples

2379-2140/19/$31.00 ©2019 IEEE
DOI 10.1109/ICDAR.2019.00059

and 400 testing samples. The dataset contains 13788 labelled text regions, where each text region represents a location of the text instance in the form of bounding boxes, the type of the script and its transcription. The Urdu-Text dataset also contains 8200-cropped words with annotations, which can be used for text recognition. To the best of authors knowledge, Urdu-Text is the first comprehensive natural scenes dataset developed for text detection and recognition. Inspired by the developments achieved for Latin text detection and recognition in natural scenes facilitated by many notable datasets, especially ICDAR's, we believe that the Urdu-Text dataset will help to extend the research work of cursive text extraction in natural scenes.


```
29,188,105,188,105,241,29,241,Urdu,بلیدی
103,197,172,197,172,239,103,239,Urdu,جنیید
170,195,229,195,229,238,170,238,Urdu,استاد
23,240,287,240,287,277,23,277,Latin,0333-7554142
300,186,576,186,576,289,300,289,Urdu,ورکشاپ
589,168,747,168,747,275,589,275,Urdu,آٹوز
745,155,1061,155,1062,313,745,313,Urdu,لاڑکانہ
514,140,665,140,656,196,514,196,Arabic,ماشاءﷲ
70,145,114,145,114,182,70,182,Latin,S
249,150,288,150,288,181,249,181,Latin,D
462,397,506,397,506,434,462,434,Arabic,ﷲ
448,435,523,435,523,469,448,469,Urdu,###
445,486,524,486,524,508,445,508,Urdu,###
445,516,524,516,524,553,445,553,Urdu,###
892,452,942,436,946,471,887,499,Urdu,###
1,568,17,568,17,730,1,730,Urdu,###
```


```
47,152,387,163,356,411,56,401,Latin,A/C
387,176,615,177,590,428,352,433,Urdu,کار
613,118,892,131,886,305,602,297,Urdu,فضیل
602,308,878,307,874,436,613,428,Urdu,رضا
14,421,132,432,121,507,16,498,Latin,915
60,515,169,520,167,597,62,596,Urdu,###
171,484,442,495,434,677,183,647,Urdu,رضا
452,485,847,486,847,670,459,665,Urdu,فضیل
185,676,432,683,430,811,185,810,Latin,A/C
438,675,613,676,615,811,442,812,Urdu,کار
629,664,726,667,724,736,623,727,Latin,###
744,654,855,657,849,729,734,730,Latin,###
335,910,469,917,473,1013,339,1001,Arabic,ﷲ
479,907,718,908,711,1016,498,1009,Arabic,ماشا
```

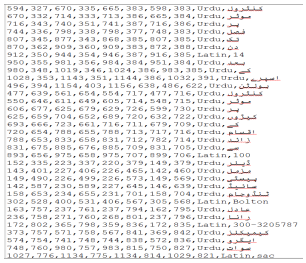

Fig. 1. Left: Some Example Images from the Urdu-Text Dataset. Right: Annotation details of Urdu-Text including rectangular bounding boxes, script type and text-transcriptions with care and don't care regions

II. Related Work

A. Natural Scene Text Datasets

In recent years, several natural scene text datasets have become popular and are used by the vast majority of the research community. These standard datasets save the efforts and time of researchers needed to compile and label the dataset. This section describes the closely related scene text datasets, some methods used to detect and recognize multi-language text in the ICDAR Multi-Language Text (MLT) dataset. Similar to other research domains, the ICDAR and other research communities have published several standard datasets to address the problem of text detection, recognition or end-to-end recognition in natural scenes and video images. The most common datasets among them are ICDAR's 03, 11, 13, 15 [1] and the 2017 Multi-lingual datasets [10], COCO-Text [4], MSRA-TD500 [2], Street View Text (SVT) [3] and Total-Text [5].

ICDAR's Robust Reading Competition: ICDAR has five variants of scene text datasets. ICDAR 03 [11] was the first dataset used to detect and recognize text in natural scenes. This dataset contains 509 natural scene images captured with a camera where the text is horizontally aligned and more iconic. The ICDAR 11 [12] dataset contains a total of 484 natural scene text images, where the text is mostly horizontally aligned. The total number of images is less than in the ICDAR 03 dataset. The next in the series of ICDAR datasets was 13, which contains 462 natural scene text images [13]. In ICDAR 15 [1], the name of the dataset is changed to the incidental scene text dataset. The number of images was also increased to 1670 where 1000 are used for training purpose, 500 are used for testing the system and the remaining 170 images are not used. This is the first of ICDAR's datasets where the images were captured without taking text into account and is considered more challenging than other scene text datasets. The most recent dataset published by ICDAR is the Multi-Lingual 2017-MLT dataset [10], which contains a total of 18000 natural scene images where 7200 are used for training, 1800 for validation and 9000 for testing purpose.

Street View Text [3] comprises 350 natural scene images taken from Google street view, which contains a total of 725 labelled words. MSRA-TD500 [2] contains a total of 500 natural scene images, which are divided into 300 training and 200 testing samples. This dataset contains text instances in multiple orientations. COCO-Text [4] is considered one of the largest natural scene text datasets containing 63683 images and 173589 labelled regions of text. The COCO-Text dataset contains text instances in all orientations: horizontal, vertical, curved and uses axis-oriented rectangles for ground truths. This dataset categorizes the text into machine printed or handwritten as well. Total-Text [5] is a recent scene text dataset, which contains multi-oriented and curved text. The dataset is divided into training and testing sets with 1255 and 300 scene images, respectively.

There are several other publicly available scene text datasets, some of them are developed for scene text recognition purposes. In [14] a synthetic dataset is presented, which contains 90000 synthetically generated cropped word images that are used to train a deep text recognition system. A dataset of scene image cropped words developed in [15] contains a total of 3000 word images.

Recently, a dataset for Arabic scene text recognition named ARASTI [16] is developed, which contains 374 scene im-

ages with Arabic text instances. This dataset describes 1280 cropped Arabic word images and 2093 manually segmented characters. However, the ARASTI dataset does not have complete image annotations with bounding boxes. Hence the ARASTI dataset can be used for Arabic scene text recognition. Related to Urdu text, a dataset of video images with artificial Urdu text is presented in [17]. This dataset contains 1000 video images taken from different Urdu news and sports channels. The ground truths are created in the form of x-coordinate, y-coordinate, width and height of bounding box rectangles. However, the artificial text does not have many variations in terms of orientations, text size, text styles, font types, patterns and background complexities. Therefore, this dataset cannot give accurate text detection results when tested for natural scene images. Real text in natural scenes, by its nature could be much more difficult to predict than artificial text. Compared to the above natural scene text recognition datasets, the Urdu-Text dataset presented in this paper contains much larger number of cropped word images, and can be evaluated for text recognition.

B. Multi-Lingual Text Detection and Urdu Character Recognition in Natural Scenes

In recent years, text detection and recognition in natural scenes has been significantly progressed with the rise of advancements in deep learning based algorithms. With the availability of ICDAR's 2017-MLT [10] and 2019-MLT [18] datasets and robust reading competition, multi-lingual scene text detection and recognition has become an active research area at the moment. Many techniques have been developed and are available in the literature. In [19] low-level and high-level features are concatenated and are shared between text detection and recognition branches. The detection branch outputs the per-pixel prediction results from the shared convolutional features and the recognition branch recognizes the text from the proposed regions using CNN and Long Short Term Memory (LSTM). The network uses ResNet-50 [20] as a backbone. In [21] segmentation and Single Short Detection (SSD) [22] networks share the same features and are combined into a single network to detect multi-oriented scene text. To detect text with multiple variances, the Feature Pyramid Network (FPN) [23] and Atrous Spatial Pyramid Pooling (ASPP) [24] are combined as encoder-decoder in a semantic segmentation branch and a new layer in an SSD network is proposed for the predictions. Some preliminary investigations on Urdu and multi-lingual character recognition in natural scene images are performed by [25]–[27].

III. Urdu-Text Dataset

This section will describe the motivation for developing the Urdu-Text dataset, text annotations, statistics and comparison of the Urdu-Text dataset with related datasets and Urdu-Text cropped word image dataset.

A. Dataset Description

To develop the Urdu-Text dataset, we photographed more than 2000 natural scene images of signboards, advertisement banners, shop names, street boards, passing vehicles and other hoardings. All the images were photographed in different cities of Sindh province, Pakistan using a digital camera with 20MP sensor, iPhone and Samsung mobile phone cameras with 12 and 16 MPs. The images in the dataset are embedded with Urdu, Latin, Sindhi and Arabic text. The statistics of the word instances for each of the language is presented in Table I. Many of the scene images are embedded with multi-lingual text. The text has variations from machine printed signboards to handwritten text on the walls as shown in Figure 1. Some of the photographed images were discarded due to un-even lighting, blur or containing no text instances. The Urdu-Text dataset contains a total of 1400 samples, where 1000 samples are annotated for the training set and 400 are selected for the testing set.

TABLE I
Statistics of the Word Instances for each Script

Script Type	No. of Words
Urdu	7603
Latin	5653
Sindhi	350
Arabic	68
Symbols	113
Others	1

B. Text Annotations

Accurate annotations of text instances is an important factor in text detection, which consequently affects the performance of the evaluation system. Some automatic or semi-automatic annotating tools have been developed, which can work correctly for Latin text, but for Urdu, Arabic or other cursive texts, these automatic tools cannot work accurately due to the non-uniform alignment of text on the same line or overlapping of text instances with other text either above or below the line as shown in Figure 2. For Arabic scene text annotation, ICDAR 2017-MLT [10] uses bounding box based annotations. Therefore, for Urdu-Text dataset, all the text regions in each image are manually annotated with a single enclosing bounding box.

Fig. 2. Scene text instances which overlap with other text

The annotations of text instances in every image are given in a separate UTF-8 encoded text file, where each line contains

the bounding box co-ordinates of a single word, the type of the text instance and text-transcription as shown in '(1)'.

$$\text{x1,y1,x2,y2,x3,y3,x4,y4, script-type, text-transcription} \quad (1)$$

where x1, y1 .... x4, y4 represent the x,y co-ordinates of the top left, top right, bottom right and bottom left corners of one word. In the Urdu-Text dataset, multi-lingual text characters including Arabic, Latin, Urdu and Sindhi scripts in natural scenes are considered. Text instances which are unreadable are considered as "don't care" regions in the ground truth file. The "don't care" areas should be filtered out when evaluating the performance of algorithms on the Urdu-Text dataset. All the mathematical signs such as +, @, /, * are annotated as a separate script-type called 'Symbol' in the annotation file. Word level annotations are also given for text recognition. Table II shows the annotations and their values collected for each text region.

TABLE II
TEXT REGION ANNOTATIONS

Annotations	Values
location	bounding box
script-type	Arabic, Latin, Sindhi, Urdu
text-transcription	String with UTF-8 encoding

For automatic annotations of Urdu text, one solution could be to identify each pixel at every location and classify either it is a text or non-text, then using some heuristic rules combine the pixels classified as text and annotate them. This process, however may require more time and effort and was not considered for this dataset.

C. Dataset Statistics and Comparison with Related Scene Text Datasets

We analysed and compared the Urdu-Text dataset with ICDAR's 2013, 2015 and 2017 multi-lingual scene text datasets [1], [10], [13]. The ICDAR's 2017-MLT dataset also contains Arabic scene text images. Therefore, the comparisons are made against Arabic scene images. The average number of text instances in Arabic scene images is 9.34, which is less than for the Urdu-Text dataset. Mostly, in Arabic scene images, the text is well focused, horizontal, machine printed and at the centre of the image. However, in the Urdu-Text dataset, the text is in multiple orientations, machine printed as well handwritten, and at various locations in the images. Compared to text patterns in the ICDAR 2017-MLT Arabic scene images, the Urdu-Text dataset has more variations in terms of text size, font type, background complexity and overlapping text as shown in Figure 3. As we can see in the second row of Figure 3, the text has very large as well as very small font sizes and it appears in multiple colors within the same image. The third and fourth images in the second row are embedded with handwritten text, which is very difficult to detect and recognize. The third row of Figure 3 shows scene images containing text in three different scripts, i.e., Latin, Urdu and Sindhi. It is also noted that the Urdu-Text dataset has more examples with handwritten text than the Arabic scene dataset, which shows that the proposed dataset has a higher level of complexity than the ICDAR 2017-MLT Arabic scene dataset.

Table III shows the statistics and comparative analysis of the Urdu-Text dataset with related datasets.

TABLE III
COMPARISON OF THE URDU-TEXT DATASET WITH RELATED DATASETS

Dataset	No. of Images	Cropped Words	No. of Text Instances	Avg. Text Instances Per Image
ICDAR 2013	462	5003	1943	4.2
ICDAR 2015	1670	6545	11886	7.12
ICDAR 2017-MLT Arabic Dataset	800	3712	7478	9.34
COCO-Text	63686	–	173589	2.73
Total-Text	1555	–	11459	7.37
ARASTI [11]	371	1687	–	–
Urdu-Text	1400	8200	13788	9.85

As shown in Table III, the Urdu-Text dataset has a larger number of cropped word images, a larger number of text instances and a larger average number of text instances per image than in the Arabic scene text dataset in ICDAR's 2017-MLT. This indicates that the proposed dataset can be used as a benchmark for reading Urdu text in natural scene images.

D. The Urdu-Text Word Image Dataset

The Urdu-Text dataset includes 8200 cropped word images with annotations in a text file. A dictionary of 40K lexicons of Urdu words is created as a text file which, can be helpful to correct the recognized words. The cropped dataset contains only the word images of Urdu and Sindhi scripts and numbers. The word instances containing Latin script are not included due to the availability of existing Latin word image datasets. Figure 4 illustrates some example images from the Urdu-Text cropped word dataset.

Fig. 4. Some Examples of Cropped Word Images

IV. EXPERIMENTS

Urdu-Text dataset is evaluated on a deep network proposed in [28], a modified deep VGG-16 network [29] with additional

Fig. 3. 1st row: Some examples from the ICDAR 2017-MLT Arabic dataset; 2nd row: Some examples from the Urdu-Text dataset; 3rd row: Some examples from the Urdu-Text dataset with multi-lingual text

layers based on pre-trained ImageNet [30] dataset and without pre-trained weights. Similar to [31] a feature fusion of convolutional layers is performed for the modified deep network, where the features at convotional level are fused together. A Fast-RCNN [32] network is applied to generate the text/non-text proposals which, are then given to the recurrent neural network for sequence to sequence connection. The network is trained using Adam optimizers at an initial learning rate of 0.001 for the initial 10K iterations, which is then decreased with an exponential decay of 0.1 after every 10K iterations. The results of the networks are evaluated using precision, recall and f-score and are shown in Table IV. Figure 5 illustrates some text detection examples of the Urdu-Text dataset. The EAST [28] text detector quite failed to detect handwritten text written on the walls as shown in the first row of Figure 5.

TABLE IV
EVALUATION OF DEEP NEURAL NETWORKS ON URDU-TEXT DATASET

Network	Precision	Recall	F-Score
EAST [28]	0.18	0.39	0.26
Custom VGG-16 [29]	0.22	0.45	0.30
Custom VGG-16 [29] with pre-trained ImageNet [30]	0.29	0.70	0.37

Fig. 5. Some text detection examples of the Urdu-Text dataset: 1st row: EAST [28]; 2nd row: Custom VGG-16 [29] without pre-trained ImageNet weights; 3rd row: Custom VGG-16 [29] with pre-trained ImageNet weights

V. CONCLUSIONS AND FUTURE DIRECTIONS

We have developed a new dataset for multi-lingual cursive script text detection and recognition, specifically focusing on Urdu text in natural scenes. This dataset will support the advancements of state-of-the-art algorithms developed for multi-lingual text detection and recognition in natural scenes. Reading text in multi-lingual scenes is a challenging task, specifically for cursive text, and research addressing this problem is in its infancy. We expect that the development of the Urdu-Text dataset presented in this paper will be a valuable resource to tackle the problem. The Urdu-Text dataset contains 1400 images, where 1000 images are manually annotated for training and 400 images are selected for testing. The dataset can be used to investigate the problem of Urdu text detection

or end-to-end text recognition in natural scene images. The dataset also contains cropped word images with labels, which could be used to recognize Urdu text in natural images. The dataset statistics show that the dataset contains a superior number of images, cropped word images and annotated text regions, which indicates that the Urdu-Text dataset can be used as a benchmark. The dataset is evaluated on three deep neural networks and the results show significant shortcomings in the accuracy, which motivates for the robust networks that can detect wide variety of text. In the future, the Urdu-Text dataset will be extended with more focused and unfocused scene text images. The text regions will also be categorised as machine printed or handwritten in annotations. The pixel level ground truths of text instances and the co-ordinates of cropped word images will also be provided.

ACKNOWLEDGEMENT

The authors are thankful to the University of New South Wales, at the Australian Defence Force Academy, Canberra, Australia for providing research and funding opportunities. The authors are also thankful to Mehwish Leghari for helping in dataset collection.

REFERENCES

[1] D. Karatzas and L. Gomez-Bigorda and A. Nicolaou and S. Ghosh and A. Bagdanov and M. Iwamura and J. Matas and L. Neumann and V. R. Chandrasekhar and S. Lu and F. Shafait and S. Uchida and E. Valveny "ICDAR 2015 competition on Robust Reading," 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, 2015, pp. 1156-1160.

[2] C. Yao, X. Bai, W. Liu, Y. Ma and Z. Tu, "Detecting texts of arbitrary orientations in natural images," 2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, 2012, pp. 1083-1090.

[3] Kai Wang, B. Babenko and S. Belongie, "End-to-end scene text recognition," 2011 International Conference on Computer Vision, Barcelona, 2011, pp. 1457-1464.

[4] A. Veit, T. Matera, L. Neumann, J. Matas, and S. Belongie. Coco-text: Dataset and benchmark for text detection and recognition in natural images. In arXiv, 2016.

[5] C. K. Ch'ng and C. S. Chan, "Total-Text: A Comprehensive Dataset for Scene Text Detection and Recognition," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 935-942.

[6] X. Yin, X. Yin, K. Huang and H. Hao, "Robust Text Detection in Natural Scene Images," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 5, pp. 970-983, May 2014.

[7] J. Ma and W. Shao and H. Ye and L. Wang and H. Wang and Y. Zheng and X. Xue, "Arbitrary-Oriented Scene Text Detection via Rotation Proposals," in IEEE Transactions on Multimedia, vol. 20, no. 11, pp. 3111-3122, Nov. 2018.

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

[9] Y. Tang and X. Wu, "Scene Text Detection and Segmentation Based on Cascaded Convolution Neural Networks," in IEEE Transactions on Image Processing, vol. 26, no. 3, pp. 1509-1520, March 2017.

[10] N. Nayef et al., "ICDAR2017 Robust Reading Challenge on Multi-Lingual Scene Text Detection and Script Identification - RRC-MLT," 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, 2017, pp. 1454-1459.

[11] S. M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong and R. Young. "ICDAR 2003 Robust Reading Competitions", In 7th International Conference on Document Analysis and Recognition - ICDAR2003, 2003.

[12] A. Shahab, F. Shafait and A. Dengel, "ICDAR 2011 Robust Reading Competition Challenge 2: Reading Text in Scene Images," 2011 International Conference on Document Analysis and Recognition, Beijing, 2011, pp. 1491-1496.

[13] D. Karatzas and F. Shafait and S. Uchida and M. Iwamura and L. G. i. Bigorda and S. R. Mestre and J. Mas and D. F. Mota and J. A. Almazàn and L. P. de las Heras, "ICDAR 2013 Robust Reading Competition," 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, 2013, pp. 1484-1493.

[14] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman. Synthetic data and artificial neural networks for natural scene text recognition. arXiv preprint arXiv:1406.2227, 2014.

[15] A. Mishra, K. Alahari, C. V. Jawahar, "Scene text recognition using higher order language priors", In BMVC-British Machine Vision Conference, pp. 1-11, 2012.

[16] M. Tounsi, I. Moalla and A. M. Alimi, "ARASTI: A database for Arabic scene text recognition," 2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR), Nancy, 2017, pp. 140-144.

[17] I. Siddiqi and A. Raza A Database of Artificial Urdu Text with Semi Automatic Text Line Labeling Scheme, In Proceedings of the 4th Int'l Conference on Advances in Multimedia(MMEDIA) Chamonix, France, pp. 75-81,2012.

[18] Rrc.cvc.uab.es. (2019). Introduction - ICDAR 2017 RobustReading Competition. [online] Available at: http://rrc.cvc.uab.es/ [Accessed 10 Feb. 2019].

[19] X. Liu, D. Liang, S. Yan, D. Chen, Y. Qiao, J. Yan, "FOTS: fast oriented text spotting with a unified network", CoRR vol. absI1801.01671, 2018.

[20] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition", Proceedings of the IEEE Conference on ComputerVision and Pattern Recognition, pp. 770-778, 2016.

[21] Y. Li, Y. Yu, Z. Li, Y. Lin, M. Xu, J. Li, and X. Zhou, Pixel-Anchor: A Fast Oriented Scene Text Detector with Combined Networks. arXiv preprint arXiv:1811.07432, 2018.

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, "SSD: Single shot multibox detector", 2015.

[23] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in Computer Vision and Pattern Recognition (CVPR), 2017.

[24] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy and A. L. Yuille, "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 834-848, 1 April 2018.

[25] A. Ali, M. Pickering and K. Shafi, "Urdu Natural Scene Character Recognition using Convolutional Neural Networks," 2018 IEEE 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), London, 2018, pp. 29-34.

[26] A. A. Chandio, M. Pickering and K. Shafi, "Character classification and recognition for Urdu texts in natural scene images," 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-6.

[27] A. Ali and M. Pickering, "Feature-Level Fusion using Convolutional Neural Network for Multi-Language Synthetic Character Recognition in Natual Images," 2018 Digital Image Computing: Techniques and Applications (DICTA), Canberra, Australia, 2018, pp. 1-6.

[28] X. Zhou, C. Yao, H. Wen, Y. Wang, S. Zhou, W. He, J. Liang. EAST: an efficient and accurate scene text detector. In Proceedings of CVPR, July 2017, pp. 2642-2651.

[29] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition. In proceedings of International Conference on Learning Representations, 2015.

[30] J. Deng, W. Dong, R. Socher, LJ. Li, K. Li, L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), July 2009, pp. 248-255.

[31] A. A. Chandio and M. Pickering, "Convolutional Feature Fusion for Multi-Language Text Detection in Natural Scene Images," 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, Pakistan, 2019, pp. 1-6.

[32] R. Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, 2015, pp. 1440-1448.