IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 21 September 2023, accepted 19 November 2023, date of publication 23 November 2023,
date of current version 6 December 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3336404

RESEARCH ARTICLE

Robust Arabic and Pashto Text Detection in Camera-Captured Documents Using Deep Learning Techniques

NISAR KHAN[1], RIAZ AHMAD[1], KHALIL ULLAH[3], SIRAJ MUHAMMAD[1], IBRAR HUSSAIN[1,2], AHMAD KHAN[4], YAZEED YASIN GHADI[5], AND HEBA G. MOHAMED[6]

[1]Department of Computer Science, Shaheed Benazir Bhutto University (SBBU), Upper Dir, Sheringal, Khyber Pakhtunkhwa 18800, Pakistan
[2]Department of Computer Science and Information Technology, University of Malakand, Chakdara, Khyber Pakhtunkhwa 18800, Pakistan
[3]Department of Software Engineering, University of Malakand (UOM), Chakdara, Khyber Pakhtunkhwa 18800, Pakistan
[4]Department of Software Engineering, Mirpur University of Science and Technology, Mirpur, Azad Jammu and Kashmir 10250, Pakistan
[5]Department of Computer Science, Al Ain University, Al Ain, United Arab Emirates
[6]Department of Electrical Engineering, Princess Nourah bint Abdulrahman University, Riyadh 11671, Saudi Arabia

Corresponding author: Ibrar Hussain (ibrar@sbbu.edu.pk)

This work was supported by Princess Nourah bint Abdulrahman University, Riyadh, Saudi Arabia, through Princess Nourah bint Abdulrahman University Researchers Supporting Project under Grant PNURSP2023TR140.

ABSTRACT In the realm of Document Image Analysis (DIA), the primary objective is to transform image data into a format that can be readily interpreted by machines. Within a DIA-based system, layout analysis plays a crucial role in pre-processing, for the identification and extraction of precise and error-free textual segments. However, regarding the Pashto language, the document images are not explored so far. Pashto text detection in camera-captured documents is a challenging task due to variations in image quality, lighting conditions, complex backgrounds unavailability of labeled documents, cursiveness, shape-context dependency, multi scripts per image, and language-specific layouts. This research examines the case of Pashto and Arabic text and contributes in two aspects. First, it introduces the creation of a real dataset that contains 1080 images of the Pashto documents captured by a handheld camera. Second, this work examines deep learning based classifiers that can perform layout analysis tasks and detects Pashto and Arabic text per document. For the layout classification, we used deep learning models such as Single-Shot Detector (SSD), Yolov5 and Yolov7. A baseline results are achieved by examining 30% images as a test set and achieve a mean average precision (mAP) of 84.51% on SSD, 88.50% on Yolov5 and 91.30% on Yolov7 respectively. The proposed methods have the potential to contribute to various applications, such as document analysis, information retrieval, and translation, for Pashto and Arabic language users.

INDEX TERMS Document image analysis, Pashto, Arabic, CNN, text detection, dataset, deep learning models.

I. INTRODUCTION

Document images are those digital images that are produced either from scanner or camera. Such documents include books, articles, postal addresses, bank cheques, forms, topographic maps, engineering drawings, license plates, and billboards, etc. [1]. These document images are in pixels form, and they could not be searched and analyzed in computers [2]. In addition to that, these images occupy large space on computer's storage and hence present a challenge to space factor. To solve and convert such kind of document images into a digital or readable format, we need a specific system where we could analyze document images. Thus, such a field is named as Document Image Analysis or shortly DIA. The DIA is the subfield of Artificial Intelligence (AI), and further, it can be categorized as one of the applications

The associate editor coordinating the review of this manuscript and approving it for publication was Turgay Celik.

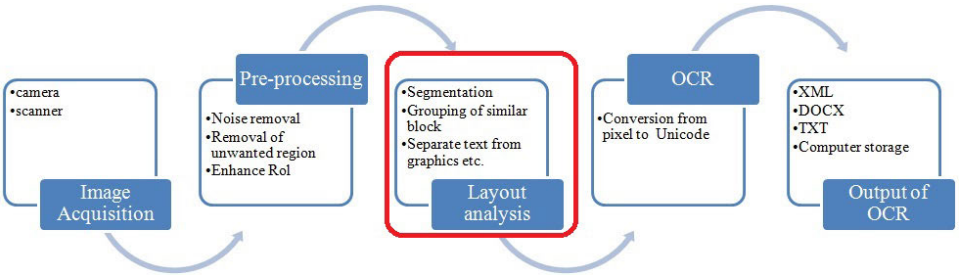


FIGURE 1. A generic view of a DIA system. The stage with a red rectangle is the focus of this article.

of the computer vision (CV). The DIA system and its five components are shown in Figure 1. While the major focus of this research is on the layout analysis and classification as highlighted in Figure 1.

Layout analysis of document and classification is a third step that is used to find out the physical structure of a document and to determine the components of a document. The fundamental components of a document are text-lines, scripts, number of columns, and non-textual regions (such as graphs, table, figures, charts, etc). Before the OCR step, these components need to be classified and their order of composition and hierarchy information need to be preserved. After the OCR stage, such preserved information is required to render back the given document [3].

This research highlights the problem of layout analysis and classification considering the Pashto language, further it analyses the very basic layouts regarding the DIA system. However, regarding the Pashto language, there is a little work in the area of DIA system. The reasons are language-specific complexities, which include; writing direction, availability of different languages per document and language-specific layouts etc. For example, several Pashto documents contain Arabic as well as Pashto text in a single document (as shown in Figure 2). Ignoring this specific pattern will lead to the extraction of textual blocks that contain either Arabic text or Pashto text or mix of both. As a result, it becomes a multi-language case for OCR, and could not be handled easily on a single OCR system.

The major contributions of this research are given below.

1) Creation of a new dataset based on camera captured images of Pashto documents.
2) Fine tuning of CNN Based deep learning models including SSD, Yolov5 and Yolov7 for baseline evaluation.

Rest of the paper is organized as follows. Section II reviews the related work. Sections III and IV introduce the dataset creation and proposed methodology. Sections V and VI present experimental setup and discuss the results in detail. Finally, Section VII discuses conclusion and future works.

II. RELATED WORK

There is an immense research work regarding DIA system. However, this research is mainly related to document layout analysis. Therefore, the related work covers only the Layout Analysis stage of a DIA system. The following section explains the very related work in detail.

A. WORK REGARDING LAYOUT ANALYSIS

O'Gorman and Kasturi [4] presented a comprehensive book that covers a document spectrum system for the structural layout analysis of the page. They have discussed various techniques including the bottom-up, nearest-clustering technique on-page elements. The book can be considered as good starting point for those who have interest in DIA systems.

Simon et al. [5] introduced a bottom-up approach for the layout analysis of document using Kruskal's algorithm [6] and to build the structure of the real page through the utilization of a particular distance metric between the segments. Their algorithm is more limited according to the computational complexity, because of its linear structure regarding the amount of the relevant elements [4].

Thomas Breuel and Warren [7] introduced many novel algorithms and statistical methods for layout analysis. The methods consist of (1) to find rectangles of tall white spaces and assess them as candidates for the channels, separators of column, etc (2) to find the text-lines concerning to the column-structure of the document, (3) to recognize paragraphs, headings and titles based on spacing, size and indentation, etc. and (4) to determine the reading order by using geometric and linguistic information. These algorithms are also applicable to Cursive Script.

Laven et al. [8] presented an algorithm using statistical patterns like grammar-based and rule-based techniques. They first introduced a unique software for the manual segmentation and labeling of the page. Their dataset contains a 932 pages as images from academic journals.[1]

Shafait et al. [9] introduced a system for the layout analysis of the cursive script. Their specified scheme experimented on 25 scanned images taken from various sources like magazines, newspapers, and books. Their algorithm obtained 90% precision in line detection, while in case of newspaper images 72% precision achieved. Shafait [10] also proposed a system for the classification and layout analysis of Breuel(Roman script text-line model) [11] to Nastaliq script

[1] http://jmlr.csail.mit.edu

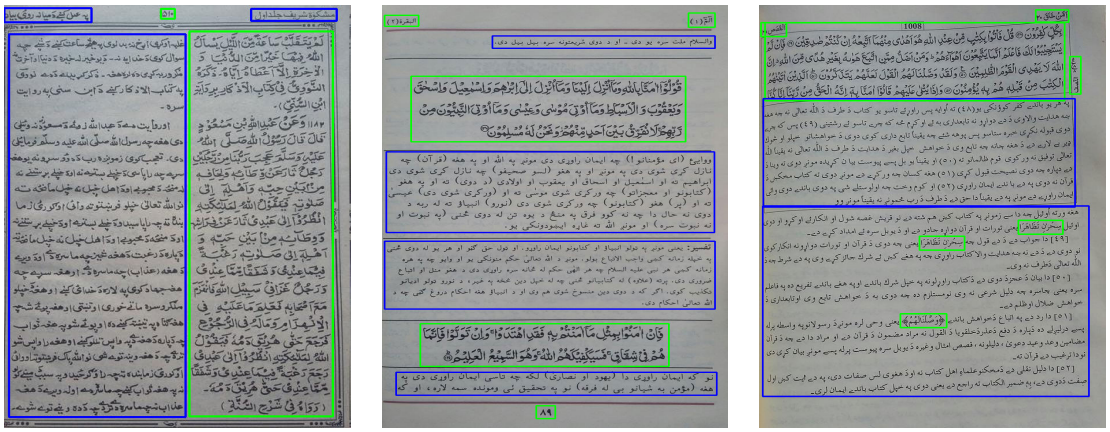


FIGURE 2. Sample images with different layouts from dataset with their ground truth annotation where Pashto text enclosed in blue bounding-box and Arabic text enclosed in light green bounding-box in a Pashto document.

and introduced a text-line extraction modeled for the Urdu text line.

Erkilinc et al. [12] introduced an algorithm for document classification and page layout analysis. They tested a module for text detection which is based on wavelet analysis and Run Length Encoding (RLE) [13] method, and a second module to detect the image and graphic sections in the input document.

Bukhari et al. [14] introduced a system for the layout analysis achieving a well organized and robust text and non-text segmentation, text-line extraction, and reading order determination methods for Urdu and Arabic document images.

Tran et al. [15] proposed a system for the analysis of the textual and non-textual elements in document images. Their method is a mixture of white-space analysis method with multi-layer uniform areas. The system was validated on page segmentation competition held by ICDAR-2009 [16]. They achieved above 90% accuracy for text detection, non-text detection, text region detection, and non-text region detection.

Ahmad et al. [17] introduced a text-line extraction method for the extraction of titles and large headings in cursive script i.e. Pashto [18], [19]. Their method is based on Horizontal Projection Pro-file (HPP) [20] and Hanning window smoothing technique [21]. They obtained an accuracy of 99.30%. However, the system needs de-skewed images for a better performance.

Goswami et al. [22] presented a multi-lingual text detector using Faster Regional Based CNN (Faster RCNN) to detect English, Hindi, and Gujarati text from Images. However, Faster RCNN is the predecessor of SSD and Yolo models. Another, their approach considers the text as one class and does not classify the text further into its respective languages. Khan and Mollah [23] used deep learning model to localized multi languages in images. However, their approach only localized text and could not recognized the text in to their respective languages.

It can be concluded that a research related to multi script detection regarding cursive script (like: Arabic and Pashto) is limited. Therefore, our research in terms of multi script detection per image document is very important and presents technological advancement to fill this gap.

III. DATASET CREATION

Data plays an important role in the evaluation of any technique. It becomes more crucial for a task like layout analysis and classification when an appropriate data is not present. Thus, this research also contributing the creation of a new dataset as a benchmark for examining layout analysis in mix languages (e.g., Arabic and Pashto) as shown in Figure 3. We call the newly dataset as Camera Captured Pashto Text Imagebase (CCPTI). Additionally, the dataset has been created from real world data that covers the real challenges of layout analysis and classification. The new dataset contains real-world samples of camera captured images presenting an appropriate benchmark for Pashto DIA system. Indeed, this research not only contributes to the scientific domain but also helps the research community regarding regional languages. On the other hand, real data presents more challenges due to the inclusion of different stages in the process. Next section explains the acquisition and creation process of a new dataset.

A. DATA ACQUISITION

The images are acquired via handheld camera. We choose two books (Tafseer-ul-Quran and Meshkat-sharif) that contain plenty of pages. However, we have captured only those pages where Pashto and Arabic text per page were notable. Additionally, while capturing the images, it was insured to avoid the skew and perspective distortion. However, blurriness and shadow effects are present due to various lighting condition in the acquired images.

B. DATASET ANNOTATION AND DESCRIPTION

In supervised learning, data must be transcribed or annotated with suitable labels. In our case, it is important to label/transcribe the Arabic and Pashto text blocks separately.

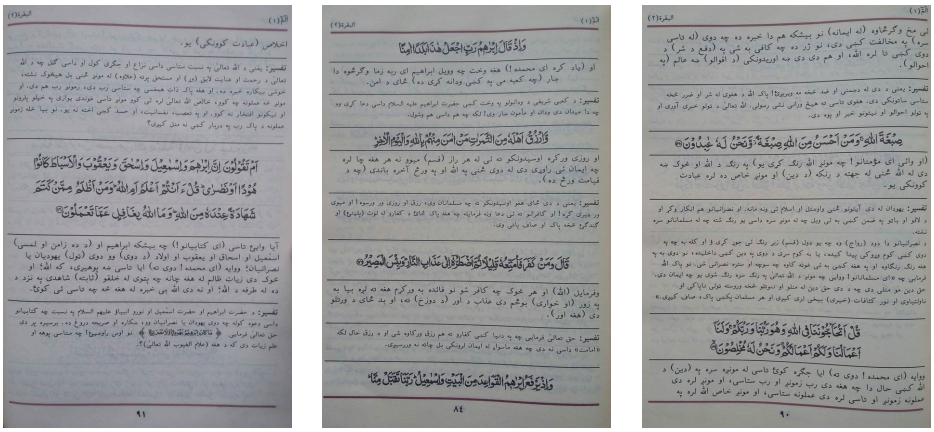


FIGURE 3. Samples of Camera-captured images containing Pashto and Arabic text-blocks.

We use a tool created by MIT named LabelMe[2] for the annotation. The annotation of each textual block is done by considering its contour/edges by taking polygons. The relevant annotation for each image is stored in a separate *.json* file. The prefix of the image filename is same as annotation or *json* file. The detailed description of the dataset is also given in Table 1. Further, in pre-processing, we did nothing except the width of each image is set to 600 pixels by keeping the aspect ratio locked. In this way, each scanned image have width as 600 pixels while its height will be of variable sizes.

TABLE 1. Sources of acquired books for our newly created CCPTI dataset.

S.No	Name of Book	Pages Captured
1	Tafseer-ul-quran	350
2	Meshkat sharif	730
	Total	**1080**

Figure 4 shows two samples from CCPTI dataset along with their ground truth illustration.

IV. PROPOSED METHODOLOGY

To accomplish the objectives and to examine the layout analysis phase of DIA in the cursive script language, we used the Convolution Neural Network(CNN) with a deep learning approach. Our proposed models are Single-shot detector(SSD) [24], Yolov5[3] and Yolov7 [25]. All the models are designed for the detection of objects by providing the bounding boxes. However, we extend these models to exploit the very basic structure of Pashto and Arabic text by assuming the textual blocks as objects. The subsequent sections explain the architecture of the proposed models in more detail.

A. SSD MODEL

Unlike its predecessor R-CNN and Faster R-CNN [26], the SSD does not use selected region proposal network, instead, it predicts the boundary boxes for the classes directly from the feature maps in a single pass. That is why it is famous for speed and performances concerning object detection problems. The very important and appealing aspect of SSD model is to handle the different scales and aspect ratios of the objects. In our case, the textual blocks may vary in size and provide different ratio with respect to height and width. Such appealing feature of SSD makes it suitable for our problem. The architecture of an SSD model consists of two main components, (1) the SSD backbone and (2) the SSD head. Here, the backbone is the Visual Geometry Group (VGG-16) network as a feature extractor [27] (similar to the CNN in Faster RCNN) [24], [26]. The SSD head usually contains one or more convolutional layers. The main function of the SSD head is to interpret the output of the SSD backbone into the bounding boxes and classes of objects in their respective spatial location via the activation of the final layers. Figure 5 shows a typical SSD model.

B. YOLOV5

Yolov5 an acronym for 'You only look once', is a complete object detection framework that divides images into a grid system. Each cell in the grid is responsible for detecting objects within itself.

C. YOLOV7

Yolov7 is the latest object detection deep learning model that super-pass all the Yolov variants with respect to accuracy and speed. Yolov7 model based on Efficient Layer Aggregation Network (ELAN) and proposed E-ELAN architecture that increase the speed and time [25].

D. EVALUATION CRITERIA

We use two different metrics to evaluate our proposed model. One is mAP(mean average precision) [28] shown in Eq: 1

$$mAP = \frac{\sum_{q=1}^{Q} avP(q)}{Q} \tag{1}$$

[2]http://labelme.csail.mit.edu/Release3.0/
[3]https://github.com/ultralytics/yolov5

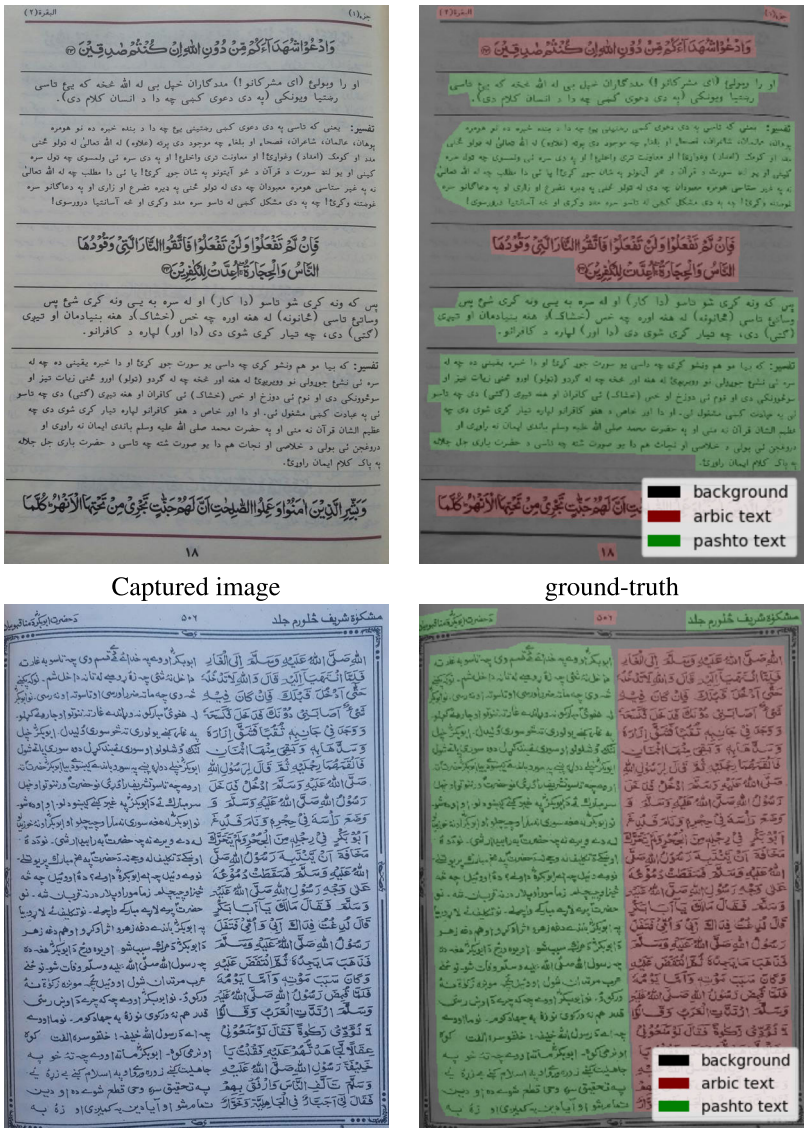


FIGURE 4. Annotated samples of Camera-captured images containing Pashto and Arabic text-blocks.

where avP(q) is the average precision (AP) for a given query and Q is the total number of queries. The second metric is IoU(intersection over union) [28] shown in Eq: 2

$$IoU = \frac{Area\ of\ Overlap}{Area\ of\ Union} \tag{2}$$

V. EXPERIMENT

The experiments were carried out by splitting the dataset into Train-set and Test-set. We used hold out method, in which nearly 70% of the data goes to training-set and the rest 30% goes to testing-set. In the training phase the Train-set was evaluated and the loss and classification error were monitored. Once the error converged, we stopped the training and the model was saved as a checkpoint. Meanwhile, the evaluation of the test-set was also done after each 100 epoch. We train three deep learning models namely SSD, Yolov5 and Yolov7.

A. EXPERIMENT ON SSD MODEL

The whole experimental process was achieved via Object Detection API [29], [30]. It is a customized tool that empowers everyone to create their powerful image classifiers. Further, the frozen model was used for the evaluation of individual as well as entire images from the test-set.

The training process has been done by Google Colab. By running Tensor-Board on our machine the training and evaluation process has also been monitored. The training ran over $20k$ steps with a batch size of 24 and achieved good results in about 8 hr 30s in 21210 epochs.

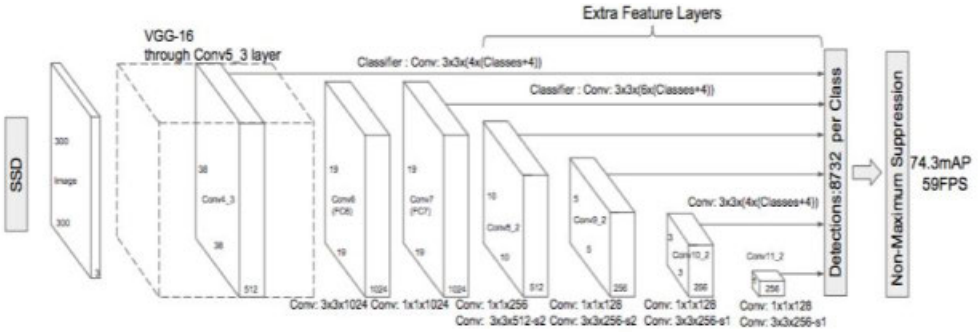


FIGURE 5. A typical Single shot detector Model (SSD) [24].

TABLE 2. Mean average precision of SSD, Detectron2, Yolov5 and Yolov7.

Model	SSD	Detectron2	Yolov5	Yolov7
mAP @ IoU =0.50	84.51%	83.61%	88.5%	91.3%

B. EXPERIMENT ON YOLOV5 MODEL

We fine-tuned 'yolov5s.pt'[4] for our dataset with image size 416, batch size 16, and number of epochs is set to 500. The training process stopped automatically after 247 epochs because the training process did not improve the accuracy for the last 100 epochs. The best accuracy is obtained at epoch No: 147.

C. EXPERIMENT OF YOLOV7 MODEL

We fin-tuned the 'yolov7.pt'[5] for our dataset with image size 640, batch size 8, and number of epochs is set to 200. The best accuracy is obtained at 62 epoch.

VI. RESULTS AND DISCUSSIONS

For SSD, after 21210 epoch, we have stopped the training, and the final model was selected for evaluation. The evaluation was done by examining the test set, and the proposed model achieved mean average precision (mAP) as 84.58%. Similarly, the same procedure was carried out to complete the training process for Yolov5 and Yolov7. The test results show that Yolov7 outperforms the SSD and Yolov5 models. TABLE 2 shows the comparison of the SSD, Yolov5 and Yolov7 models in terms of mAP.

Further, to check the effectiveness of the proposed work, a counter experiment has been done on Detectron2[6] [31] model. The Detectron2 model is famous for its effectiveness in Facebook platform and its better performance on PubLayNet[7] Dataset [32], [33]. The Detectron2 was fine-tuned on our dataset for 3000 epochs and then the final model was evaluated on the test set. The overall performance of the models that we evaluated plus the comparison on the Detectron2 model is given in TABLE 2.

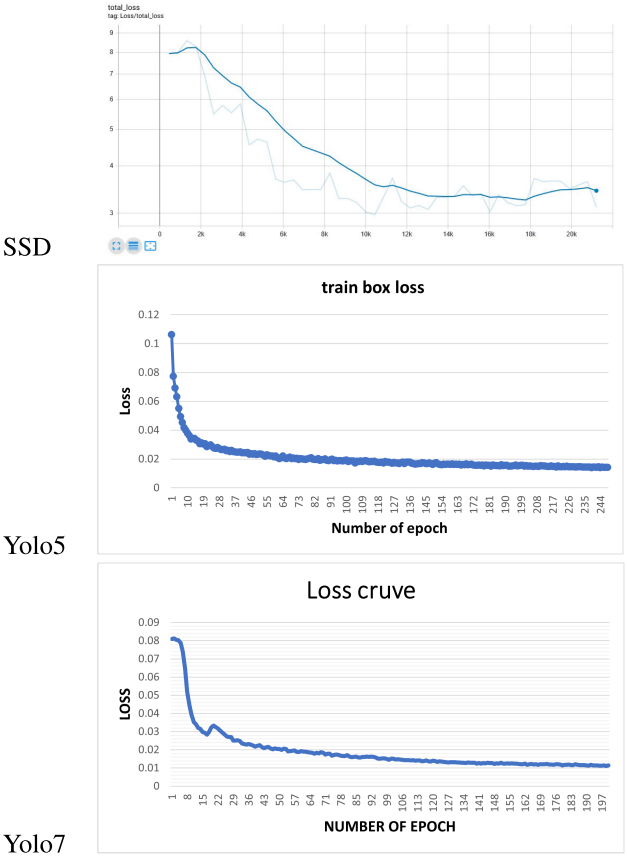


FIGURE 6. The loss was gradually reduced for SSD, yolo5 and yolo7.

Similarly, the training processes were tested for how they reduce the total loss for SSD, Yolov5 and Yolov7 models. Figure 6 depicts the value of loss regarding each epoch in the training phase for the respective models.

A. DISCUSSION

After a detailed analysis of results as well as individual images, we have findings that are explained in the following sections. Before we could empirically assess the

[4]https://github.com/ultralytics/yolov5/releases/download/v6.2/yolov5s.pt
[5]https://github.com/WongKinYiu/yolov7/releases/download/v0.1/yolov7.pt
[6]https://github.com/facebookresearch/detectron2
[7]https://paperswithcode.com/dataset/publaynet

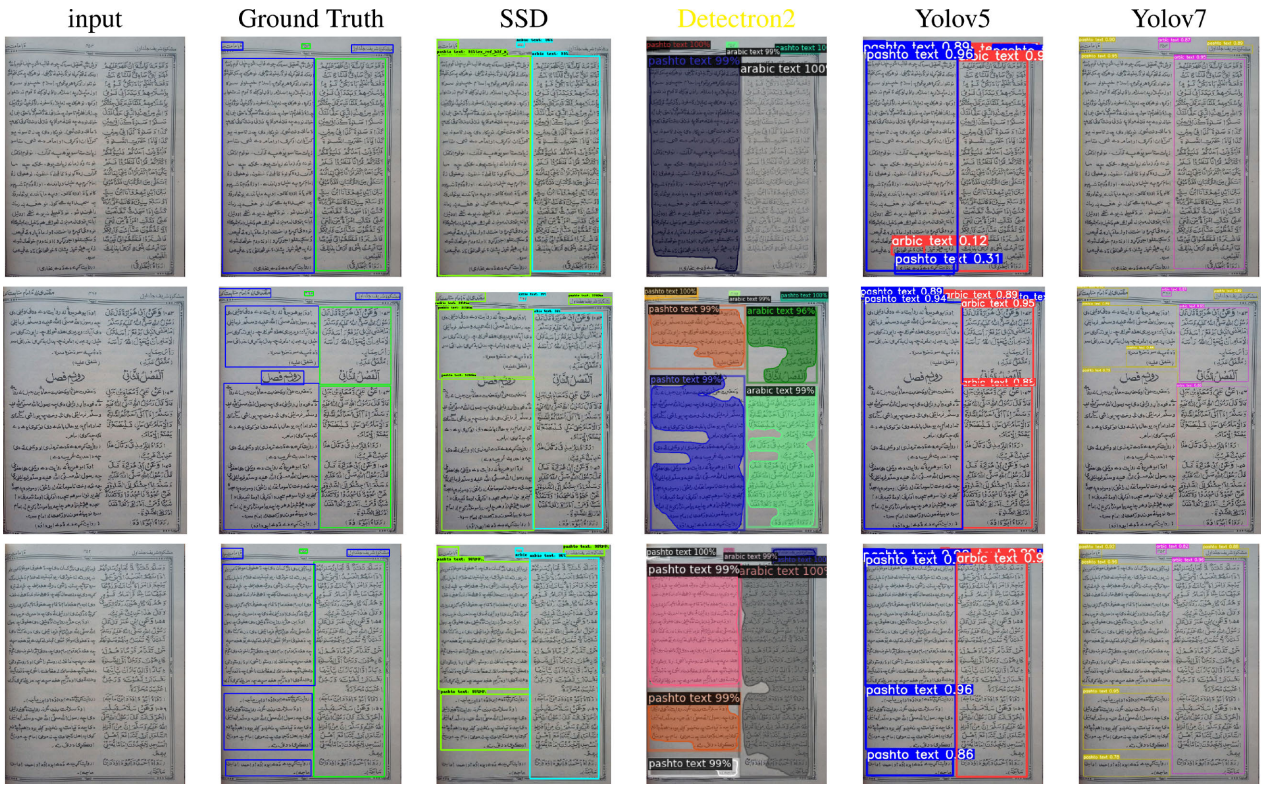


FIGURE 7. Visual results for SSD, Yolov5 and Yolov7 compared to the original input image along with ground truth.

results, Figure 7 shows some prediction of our proposed models on completely unseen data. To understand the visual representation of our examples, please understand the notions mentioned here. The blue, yellow, green color for the bounding boxes represents "Pashto text" and the red, magenta, aqua and light green color represent "Arabic text" while GT represents the bounding boxes for the ground truth.

B. OBSERVATION NO 1

The first row in the Figure 7 represents the case where SSD, Yolov5 and Yolov7 models have shown similar performance and produces equal detection results.

C. OBSERVATION NO 2

The middle row represents images with a moderate complexity which is shown in Figure 7. Visually all the three models have given acceptable detection of textual blocks. However, in terms of bounding boxes, some large text with smaller bounding boxes were skipped. But, that small bounding box was enclosed in a larger textual block. Apparently, this should not be an error but regarding the IoU metric, it leads to increase the error. Besides these errors, still all the three models performed well and equally.

D. OBSERVATION NO 3

However, the last row in Figure 7 shows some images with fewer detection and miss-classification for SSD model. However, Yolov5 and Yolov7 show the better results than SSD. Visual inspection yields, that SSD was not that much effective to classify textual blocks that are enclosed in smaller bounding boxes. On other hand, the Yolov5 and Yolov7 performed better in classification of textual blocks that are having small bounding boxes.

E. OBSERVATION NO 4

With respect to time comparison, Yolov7 produce better results than Yolov5 and SSD model. Yolov7 converge and produce best results at epoch 62, Yolov5 produce at epoch 147 and SSD produce the best result at 21210. It is due to the fact that Yolov7 architecture is reform with new module such as Extended Efficient Layer Aggregation Network (E-ELAN).

VII. CONCLUSION AND FUTURE WORK

This work for the first time presents a study regarding the layout analysis and classification of Pashto document images. The research particularly examined the classification of Arabic text vs Pashto text in Pashto document images. This work contributes mainly in two aspects. First, we have created a new dataset that contains real Pashto document images. The images are acquired via a handheld camera. The dataset will be a significant resource for the research community for analysing the DIA domain in cursive scripts.

Further, the second contribution is the application of deep learning-based methods to examine how we can detect/classify Arabic text vs Pashto text in a single document image. We have chosen the SSD, Yolov5 and Yolov7 models. The SSD, a hybrid model containing VGG16 as convolutional layers and a Neural network for learning high distinctive features. On other hand, Yolo5 and Yolo7 are famous for their light weight, speed and accuracy. Yolov7 extends the Efficient Layer Architecture Network known as E-ELAN which increases the accuracy and speed of Yolov7 among all Yolo versions. We achieved an mAP of 84.51% on SSD model, 88.50% mAP on Yolov5, and 91.30% mAP on Yolov7. The results show that Yolov7 is far better than Yolov5 and SSD. In addition to deep learning models like, SSD, Yolov5, and Yolov7, a comparison was also done with Detectron2 model by fine-tuning the already trained model on our train set. However, evaluation on test set gives us mAP as 83.61% which is comparatively less than Yolov5 and Yolov7.

In future, we would like to extend the dataset with more layouts that are present in the contents of the Pashto language. Also, we can used CNN based architectures that are more sophisticated compared to the SSD, Yolov5 and Yolov7.

REFERENCES

[1] K. D. Kalaskar and M. P. Dhore, ''Preprocessing challenges in document image analysis,'' in *Proc. MPGI Nat. Multi Conf. (MPGINMC*, 2012, pp. 1–4.

[2] U. D. Dixit and M. Shirdhonkar, ''A survey on document image analysis and retrieval system,'' *Int. J. Cybern. Informat.*, vol. 4, no. 2, pp. 259–270, Apr. 2015.

[3] R. Kumar, M. Mukerji, and H. Gur, ''A fast multifunctional approach for document image analysis,'' in *Proc. 7th Int. Conf. Document Anal. Recognit.*, Aug. 2005, p. 1178.

[4] L. O'Gorman and R. Kasturi, *Document Image Analysis*, vol. 39. Los Alamitos, CA, USA: IEEE Computer Society Press, 1995.

[5] A. Simon, J.-C. Pret, and A. P. Johnson, ''A fast algorithm for bottom-up document layout analysis,'' *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 3, pp. 273–277, Mar. 1997.

[6] P. B. Guttoski, M. S. Sunye, and F. Silva, ''Kruskal's algorithm for query tree optimization,'' in *Proc. 11th Int. Database Eng. Appl. Symp. (IDEAS)*, Sep. 2007, pp. 296–302.

[7] I. Bruegel and S. Warren, ''Family resources and community social capital as routes to valued employment in the U.K.?'' *Social Policy Soc.*, vol. 2, no. 4, pp. 319–328, Oct. 2003.

[8] K. Laven, S. Leishman, and S. Roweis, ''A statistical learning approach to document image analysis,'' in *Proc. 8th Int. Conf. Document Anal. Recognit. (ICDAR)*, Aug. 2005, pp. 357–361.

[9] F. Shafait, Adnan-ul-Hasan, D. Keysers, and T. M. Breuel, ''Layout analysis of Urdu document images,'' in *Proc. IEEE Int. Multitopic Conf.*, Dec. 2006, pp. 293–298.

[10] F. Shafait, ''Geometric layout analysis of scanned documents,'' Ph.D. dis-sertation, Dept. Comput. Sci., Univ. Kaiserslautern, Kaiserslautern, Germany, 2008.

[11] T. M. Breuel, ''Robust least-square-baseline finding using a branch and bound algorithm,'' *Proc. SPIE*, vol. 4670, pp. 20–28, Dec. 2001.

[12] M. S. Erkilinc, M. Jaber, E. Saber, P. Bauer, and D. Depalov, ''Page layout analysis and classification for complex scanned documents,'' *Proc. SPIE*, vol. 8135, Sep. 2011, Art. no. 813507.

[13] S. C. Hinds, J. L. Fisher, and D. P. D'Amato, ''A document skew detection method using run-length encoding and the Hough transform,'' in *Proc. 10th Int. Conf. Pattern Recognit.*, Jun. 1990, pp. 464–468.

[14] S. S. Bukhari, F. Shafait, and T. M. Breuel, ''Layout analysis of Arabic script documents,'' in *Guide to OCR for Arabic Scripts*. Cham, Switzerland: Springer, 2012, pp. 35–53.

[15] T.-A. Tran, I.-S. Na, and S.-H. Kim, ''Separation of text and non-text in document layout analysis using a recursive filter,'' *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 10, pp. 1–20, 2015.

[16] A. Antonacopoulos, S. Pletschacher, D. Bridson, and C. Papadopoulos, ''ICDAR 2009 page segmentation competition,'' in *Proc. 10th Int. Conf. Document Anal. Recognit.*, Jul. 2009, pp. 1370–1374.

[17] R. Ahmad, M. Z. Afzal, S. F. Rashid, M. Liwicki, and A. Dengel, ''Text-line segmentation of large titles and headings in Arabic like script,'' in *Proc. 1st Int. Workshop Arabic Script Anal. Recognit. (ASAR)*, Apr. 2017, pp. 168–172.

[18] R. Ahmad, M. Z. Afzal, S. F. Rashid, M. Liwicki, T. Breuel, and A. Dengel, ''KPTI: Katib's pashto text imagebase and deep learning benchmark,'' in *Proc. 15th Int. Conf. Frontiers Handwriting Recognit. (ICFHR)*, Oct. 2016, pp. 453–458.

[19] R. Ahmad, S. Naz, M. Z. Afzal, S. F. Rashid, M. Liwicki, and A. Dengel, ''The impact of visual similarities of Arabic-like scripts regarding learning in an OCR system,'' in *Proc. 14th IAPR Int. Conf. Document Anal. Recognit. (ICDAR)*, vol. 7, Nov. 2017, pp. 15–19.

[20] M. Javed, P. Nagabhushan, and B. B. Chaudhuri, ''Extraction of projection profile, run-histogram and entropy features straight from run-length compressed text-documents,'' 2014, *arXiv:1404.0627*.

[21] A. Testa, D. Gallo, and R. Langella, ''On the processing of harmonics and interharmonics: Using Hanning window in standard framework,'' *IEEE Trans. Power Del.*, vol. 19, no. 1, pp. 28–34, Jan. 2004.

[22] M. M. Goswami, N. J. Dadiya, S. Mitra, and T. Goswami, ''Multi-script text detection from image using FRCNN,'' *Int. J. Asian Lang. Process.*, vol. 31, no. 2, Jun. 2021, Art. no. 2250003.

[23] T. Khan and A. F. Mollah, ''A novel multi-scale deep neural framework for script invariant text detection,'' *Neural Process. Lett.*, vol. 54, no. 2, pp. 1371–1397, Apr. 2022.

[24] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, ''SSD: Single shot multibox detector,'' in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2016, pp. 21–37.

[25] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, ''YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors,'' 2022, *arXiv:2207.02696*.

[26] R. Girshick, ''Fast R-CNN,'' in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1440–1448.

[27] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' 2014, *arXiv:1409.1556*.

[28] M. Everingham and J. Winn, ''The PASCAL visual object classes challenge 2012 (VOC2012) development kit,'' Pattern Anal., Stat. Model. Comput. Learn., Oxford Univ., U.K., Tech. Rep., 2011.

[29] J. Huang, V. Rathod, D. Chow, C. Sun, M. Zhu, A. Fathi, and Z. Lu. (2017). *Tensorflow Object Detection API*. [Online]. Available: Code:github.com/tensorflow/models/tree/master/object detection

[30] J. Huang, V. Rathod, C. Sun, M. Zhu, A. Korattikara, A. Fathi, I. Fischer, Z. Wojna, Y. Song, S. Guadarrama, and K. Murphy, ''Speed/accuracy trade-offs for modern convolutional object detectors,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3296–3297.

[31] A. V. S. Abhishek and S. Kotni, ''Detectron2 object detection & manipulating images using cartoonization,'' *Int. J. Eng. Res. Technol. (IJERT)*, vol. 10, pp. 1–5, Aug. 2021.

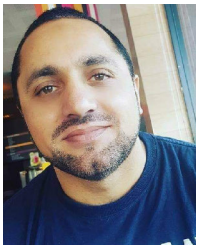
[32] A. B. Abdusalomov, B. M. S. Islam, R. Nasimov, M. Mukhiddinov, and T. K. Whangbo, ''An improved forest fire detection method based on the Detectron2 model and a deep learning approach,'' *Sensors*, vol. 23, no. 3, p. 1512, Jan. 2023.

[33] F. J. Yagüe, J. F. Diez-Pastor, P. Latorre-Carmona, and C. I. G. Osorio, ''Defect detection and segmentation in X-ray images of magnesium alloy castings using the Detectron2 framework,'' 2022, *arXiv:2202.13945*.

NISAR KHAN received the BCS degree (Hons.) from the Department of Computer Science, University of Malakand, Pakistan, and the M.S. degree in computer science from the Department of Computer Science, Shaheed Benazir Bhutto University, Upper Dir, Sheringal.

RIAZ AHMAD received the M.S. degree (Hons.) in computer science from NUCES (FAST) University, Pakistan, in 2010, and the Ph.D. degree from the Technical University of Kaiserslautern, Germany, in 2018. He was a member with the Multimedia Analysis and Data Mining (MADM) Research Group, German Research Center for Artificial Intelligence (DFKI), Kaiserslautern, Germany. He is currently heading the Computer Science Department, Shaheed Benazir Bhutto University, Sheringal, Pakistan. His research interests include document image analysis, image processing, and optical character recognition. More specifically, his work examines the challenges posed by cursive script languages in the field of OCR systems. In addition to that, he is studying the behavior of deep learning architectures in the field of OCR in terms of invariant approaches against scale and rotation variation in Pashto cursive text.

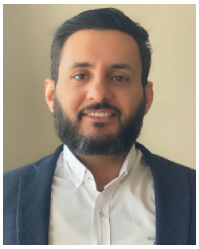
KHALIL ULLAH received the degree in computer systems engineering from the University of Engineering and Technology, Peshawar, Pakistan, in 2006, the Master of Science (M.S.) degree in electronics and communications engineering from Myongji University, South Korea, in 2009, and the Ph.D. degree in biomedical engineering from LISiN, Politecnico di Torino, in 2016, under Erasmus Mundus Expert II fellowship. He is currently acting as an Assistant Professor and the Head of the Software Engineering Department, University of Malakand. His research interests include extracting muscle anatomical and physiological information from high-density electromyography, computer vision, digital signal and image processing, and deep learning with applications to medical healthcare.



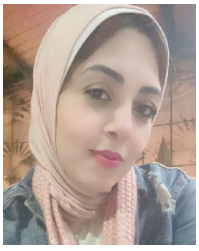
SIRAJ MUHAMMAD received the M.Phil. degree from Quaid-i-Azam University, Islamabad, in 2010, and the Ph.D. degree from the Asian Institute of Technology (AIT), Thailand, in 2020. He was a Software Engineer with Elixir Technologies, Islamabad, Pakistan, from 2010 to 2011. He is currently an Assistant Professor with the Department of Computer Science, Shaheed Benazir Bhutto University, Sheringal, Pakistan. His research interests include reverse engineering, computer vision, image processing, deep learning, and natural language processing.

IBRAR HUSSAIN received the BCS degree (Hons.) from the Department of Computer Science, University of Peshawar, Pakistan, and the M.S. degree in computer science from the Department of Computer Science, Shaheed Benazir Bhutto University, Upper Dir, Sheringal. He is currently pursuing the Ph.D. degree with the Department of Computer Science and Information Technology, University of Malakand, Lower Dir, Chakdara, Khyber Pakhtunkhwa, Pakistan.

AHMAD KHAN received the B.Sc. degree in computer systems engineering, the M.Sc. degree in computer software engineering, and the Ph.D. degree in computer systems engineering from the University of Engineering and Technology (UET), Peshawar, Pakistan, in 2006, 2014, and October 2020, respectively. He is currently an Assistant Professor of software engineering with the Mirpur University of Science and Technology, Mirpur, Azad Jammu and Kashmir, Pakistan. His research interests include machine to machine communication, the Internet of Things (IoT), and computer vision.



YAZEED YASIN GHADI received the Ph.D. degree in electrical and computer engineering from Queensland University. He is currently an Assistant Professor of software engineering with Al Ain University. He was a Postdoctoral Researcher with Queensland University, before joining Al Ain University. He has published more than 80 peer-reviewed journal and conference papers and holds three pending patents. His current research interests include developing novel electro-acousto-optic neural interfaces for large scale high resolution electrophysiology and distributed optogenetic stimulation. He was a recipient of several awards. His dissertation on developing novel hybrid plasmonicphotonic onchip biochemical sensors received the Sigma Xi Best Ph.D. Thesis Award.



HEBA G. MOHAMED was born in Alexandria, Egypt, in 1984. She received the B.Sc. and M.Sc. degrees in electrical engineering from the Arab Academy for Science and Technology, in 2007 and 2012, respectively, and the Ph.D. degree in electrical engineering from the University of Alexandria, Egypt, in 2016. In 2016, she was an Assistant Professor with the Alexandria Higher Institute of Engineering and Technology, Ministry of Higher Education, Egypt. Since 2019, she has been an Assistant Professor with the Faculty of Engineering, Communication Department, Princess Nourah bint Abdulrahman University, Saudi Arabia. In 2022, she became an Associate Professor in Egypt. Her research interests include cryptography, wireless communication, mobile data communication, the Internet of Things, and computer vision.