Arabic Words Extraction and Character Recognition from Picturesque Image Macros with Enhanced VGG-16 based Model Functionality Using Neural Networks

Ayed Ahmad Hamdan Al-Radaideh[1], Mohd Shafry bin Mohd Rahim[1], Wad Ghaban[2], Majdi Bsoul[3], Shahid Kamal[4,*], and Naveed Abbas[5]

[1]School of Computing, Universiti Teknologi Malaysia, Johor Bahru, Malaysia
[2]Applied College, University of Tabuk, Tabuk, 47512, Saudi Arabia
[3]School of Engineering, Swansea University, Swansea, Wales SA2 8PP UK
[4]Self-employed computer science researcher, D.I.Khan, KP, Pakistan
[e-mail : skamaltipu@gmail.com]
[5]Department of Computer Science, Islamia College Peshawar, KP, Pakistan
[*]Corresponding author: Shahid Kamal

Received March 21, 2023; revised May 16, 2023; accepted July 3, 2023; published July 31, 2023

Abstract

Innovation and rapid increased functionality in user friendly smartphones has encouraged shutterbugs to have picturesque image macros while in work environment or during travel. Formal signboards are placed with marketing objectives and are enriched with text for attracting people. Extracting and recognition of the text from natural images is an emerging research issue and needs consideration. When compared to conventional optical character recognition (OCR), the complex background, implicit noise, lighting, and orientation of these scenic text photos make this problem more difficult. Arabic language text scene extraction and recognition adds a number of complications and difficulties. The method described in this paper uses a two-phase methodology to extract Arabic text and word boundaries awareness from scenic images with varying text orientations. The first stage uses a convolution auto-encoder, and the second uses Arabic Character Segmentation (ACS), which is followed by traditional two-layer neural networks for recognition. This study presents the way that how can an Arabic training and synthetic dataset be created for exemplify the superimposed text in different scene images. For this purpose a dataset of size 10K of cropped images has been created in the detection phase wherein Arabic text was found and 127k Arabic character dataset for the recognition phase. The phase-1 labels were generated from an Arabic corpus of quotes and sentences, which consists of 15kquotes and sentences. This study ensures that Arabic Word Awareness Region Detection (AWARD) approach with high flexibility in identifying complex Arabic text scene images, such as texts that are arbitrarily oriented, curved, or deformed, is used to detect these texts. Our research after experimentations shows that the

http://doi.org/10.3837/tiis.2023.07.004 **ISSN : 1976-7277**

system has a 91.8% word segmentation accuracy and a 94.2% character recognition accuracy. We believe in the future that the researchers will excel in the field of image processing while treating text images to improve or reduce noise by processing scene images in any language by enhancing the functionality of VGG-16 based model using Neural Networks.

Keywords: Arabic Characters Recognition, AWARD, Deep learning, Scenic Images Dataset, VGG-16, Neural Networks.

1. Introduction

In the recent digitization era, wherein smartphones equipped with increasing functionality have enable the users to take sight photographs in order to use the embedded text messages so as to extract latest information related with several advancements in different products. In order to detect and recognize Arabic words from these photographs, Arabic Word Awareness Region Detection (AWARD) framework is applied in this study. In Past, ASCII characters or Optical Character Reader (OCR) were used to classify the patterns superimposed on images by converting scanned documents into different formats and then extract text by using editable design. Which was then applied to detect Arabic words, known as Arabic Optical Character Recognition (AOCR) [1]. The accuracy of the approaches may not be suitable for cursive languages like Arabic, because neural network based text detection and character recognition techniques in scenic images are primarily developed for English language and are often based on word-level bounding boxes. In the cases of cursive languages like Arabic, boundaries between characters and words are less well-defined and the text may be more irregular in shape. Complex backgrounds in the photographs have made it more difficult to resolve the issue when compared to traditional OCR because of noise, ambient lighting conditions, etc.

Text detection from scene photos can be done using a variety of applications, including those for immediate translation of text, image as well as information retrieval, and Electronic Orientation Aids (EOAs), etc. In more than 25 countries of the world, Arabic has been recognized an as official language. Furthermore, its vocabulary is also part of other spoken languages in different counties as well with population more than 600,000,000 [2-5]. In the Arabic style wherein text is written in the right to left pattern, causes different changes in the shape of the characters in a word associated with its position within a word. In AOCR research, an interesting area is segmentation wherein words are divided into different sub-words as well as into their individual characters. In Past, word level bounding boxes were used for this purpose, but these methods have been abandoned due to unexpected results when applied in the case of cursive nature languages like Arabic in this study.

In contrast to the word level awareness and decomposing the word into individual characters for recognition has plenty merits. In any research area, the fundamental requirement is the existence of dataset for the smooth conduction of research. In order to cope with handwritten text issues, several datasets include image macros enriched with handwritten samples. However, these image macros need to be evaluated and tested with large scale systems using advanced practices. In this study as contribution, we developed an Arabic training and testing dataset comprising scenic photos that have Arabic text superimposed on them. 10,000 cropped photos with Arabic text contained make up the dataset. An Arabic word corpus of 15,000

words was used to create the labels. Noise reduction and background removal have been done during the preparation of the scene photos. The related text-scene images are then given the ground truth annotations.

In order to detect text area and then generate region as well as affinity scores between Arabic characters, Arabic Optical Characters Recognition (AOCR) along with online or offline characteristics is used for extracting the features to identify Arabic word boundary. For assigning score, each pixel in the Arabic characters has been used for creating geometric boundary boxes. Afterward segmentation, character affinity scores [6] are applied.

After detecting text area amongst Arabic characters and generating their region as well as affinity scores by using features extraction methods for identifying Arabic word boundaries, these characters are pass in to conventional Neural Network with two hidden layers (NN-2L) model for the purpose to effectively recognize the Arabic characters. Such that an Arabic word may be composed.

The two phases of AWARD methodology is compared with classical OCR methods depicted in **Fig. 1** below. This research study will contribute in the field of image processing in terms of creating a new dataset of 10,000 cropped images, which can be enhanced further for more efficient results; similarly, it provides a way of learning to extract text from scene images. In this study, we have limited our scope only in the Arabic language which can further be applied to any language of the world.

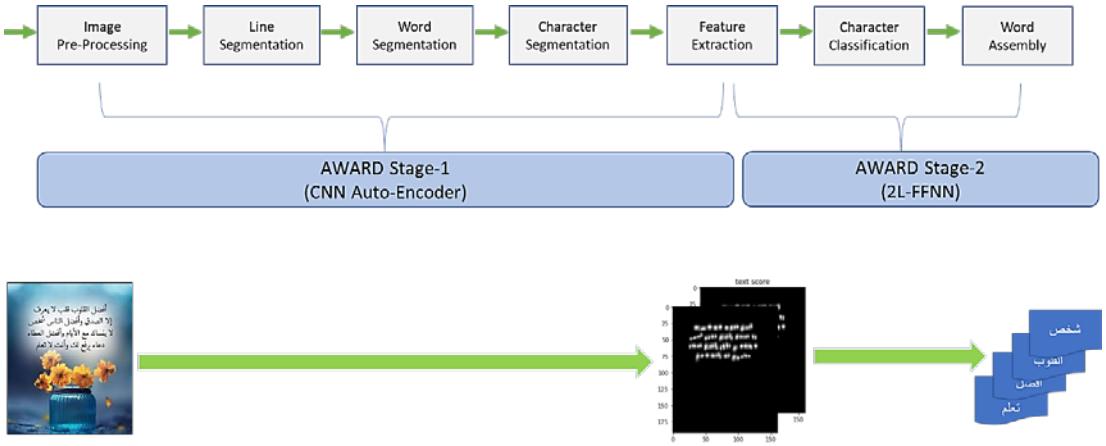

Fig. 1. Classical OCR pipeline vs AWARD Methodology

2. Related Work

Since last three decades, extensive research has been carried out in other languages likewise Chinese but as Arabic language is more known as semantic language and is widely used; therefore it needs utmost consideration in domain of character recognition [7]. The researchers in 1980 put focus on extraction of Arabic words for different purposes likewise segmentation etc. and achieved remarkable improvement in it. There is a method called as Arabic Optical Character Recognition (AOCR) wherein scanned documents containing images are converted into PDF or ASCII formats which are editable using computer system. In the same way, alphanumeric characters are classified into different patterns and this process is called as Optical Character Recognition (OCR). Classifying optical patterns in digital images, including text images, requires the use of image processing algorithms. The processes used in these techniques, which include image enhancement, filtering, segmentation, and feature extraction,

can enhance the image's quality and speed up further processing steps like classification and recognition. Additionally, Machine Learning (ML) techniques are increasingly in demand for jobs involving the categorization and recognition of images, notably in the area of medical imaging. In order to uncover patterns and characteristics that are pertinent for classification and recognition, ML algorithms can learn from massive datasets of labelled images. This process calls for a combination of image processing techniques as well as the requisite prior knowledge and skills to use them successfully [8].

Our ability to recognize and classify objects, people, and events based on their traits and behaviors is a fundamental cognitive process. Recognition is crucial for human cognition, but it also has useful applications in Computer Vision, Image Processing, Natural Language Processing (NLP), and Machine Learning (ML). Recognition algorithms are employed in various domains to recognize and categorize objects, patterns, and data in digital photographs and other types of data. An object has features or attributes which describe it and their combination defines pattern of the object. Correspondingly, many objects share the same pattern fall in a same class known as the pattern class [1].

The classification or recognition of a pattern is either supervised or unsupervised and can be applied in identifying fingerprint, voice, face, signature, and character recognition. Besides its usage in medical and ML research, it can also be used in different fields likewise mathematics, astronomy etc. however, in general it consists of three basic components namely; data preprocessing, data representation, and decision making [9].

Utilizing Arabic printed text recognition system in historical and ancient Arabic manuscripts at least more than three million can enable the automatic searching and reading [10]. Since Arabic printing hasn't changed much throughout time, Modern Standard Arabic has adopted the same methods. But, while considering those manuscripts for image degradation cause unexpected markings that fallouts into enormously confusing the recognition task [4].

The character recognition has attracted the researchers since last three decades to make significant contributions in the area of pattern recognition. But Arabic language remained unfocused while comparing research in other languages of the world [5]. In the context to recognize Arabic language [7] has made first attempt, while in 1940 [11] the authors have claimed their attempt for Latin recognition system. In addition to all these the scenic text has much more challenges especially in the context of Arabic language. While considering these mentioned research gaps, we have focused upon Arabic character as well as word recognition in combine by using scene text imaging to simulate human reading capabilities linked with activities defined in [12, 13]. While considering research in the pattern recognition filed, languages remained well thought out [14]. Previously, Arabic character recognition dealt with two domains namely WR and CR; but now scenic text has significantly attracted the researchers and is considered as third domain in PR. At first, scanned text images are transferred into digitized format and then at second step, by using digitizer devices likewise digital pen etc., handwritten text is captured.

The advent of deep learning has revolutionized the approach to scenic text detection. Prior to deep learning, bottom-up approaches to text detection were commonly used. These methods relied on low-level features such as edge detection, texture analysis, and color information to identify text regions in an image. However, these methods were often unreliable and prone to errors, particularly in cases where the text was distorted, occluded, or written in a non-standard font. Therefore, the use of deep learning in scenic text detection has led to significant improvements in accuracy and reliability, and has opened up new possibilities for applications in fields such as OCR, NLP, computer vision, MSER [14] and SWT [4] wherein handcrafted

features were used as foundation. In fact, the rise of deep learning has sparked the creation of fresh approaches to object segmentation, including the Single Shot Detector (SSD) [15-17]. Using a single network, SSD combines object localization and classification to create an object detection approach based on deep learning. Each object's bounding boxes and labels are predicted using a neural network, and semantic segmentation is used to separate the object from its background. Most scenic text detectors use box regression techniques to detect and localize text regions in an image. However, one of the main challenges with scenic text detection is the irregular shapes of text due to various factors such as perspective distortion, curved surfaces, and overlapping with other objects in the scene while Textboxes [18] is a method that addresses this challenge by using convolutional kernels to bound the text shapes according to their appearance. Instead of relying on rectangular bounding boxes, Textboxes uses oriented bounding boxes that can better fit the irregular shapes of text. These bounding boxes are defined by a set of parameters that describe their location, size, and orientation. The Deep Matching Network (DMNet) [19] uses quadrilateral sliding windows as another method to deal with the issue of text in scenic photos with irregular shapes. Using sliding windows with a quadrilateral form, the quadrilateral slide windows technique locates text sections in images. When the text sections have intricate shapes that cannot be correctly represented by rectangular bounding boxes, this method is especially helpful. By applying rotation filters to create rotation-invariant features, the Rotation-Sensitive Scenic Detection (RSSD) [20] approach presented a solution to the issue of the cursive nature of scenic text. Cursive text poses a significant challenge for scenic text detection methods because the shapes and orientations of the characters can vary significantly within a single word or sentence. This makes it difficult to use traditional bounding box-based approaches for text detection. In scenic text detection, segmentation-based methods have been applied to separate text sections at the pixel level. These techniques forecast a segmentation map using Convolutional Neural Networks (CNNs), where each pixel is given a label designating whether or not it is a text region. The MS-FCN [4], a segmentation-based approach is a deep neural network design that extracts features from the input image at various scales using many layers of convolutional and pooling layers. The network then creates a segmentation map that pinpoints text sections at the pixel level using a decoder network. The Synthetically Supervised Scene Text Detector (SSTD) [21] approach trains a deep neural network to recognize text regions in real-world photos using a synthetic dataset of text images and background images. The network is trained to anticipate the text regions' bounding box and segmentation mask during training. The advantages of segmentation- and regression-based techniques are combined in the SSTD method. While the segmentation-based technique enables the network to recognize the text region's pixel-level borders, the regression-based approach enables the network to forecast the text region's exact location. The TextSnake method proposed by Long et al. in 2018 [22] is a more contemporary model for text recognition in scenic photos that predicts text locations using geometric properties. It employs a deep neural network to forecast each polygon segment's coordinates and orientation as well as the likelihood that each segment is a part of a text region. The Fused Text Spotting (FOTS) [23] and EAA [24] are two current techniques that combine the detection and recognition processes in an end-to-end manner to achieve high accuracy in text detection and recognition. In order to predict both the text region and the associated text string within the region, the TextSpotter [25] approach employs a deep neural network. The approach combines a semantic segmentation network for character recognition with a fully convolutional network for text region proposal. In order to forecast the class of each pixel in the text region, including the characters, background, and other non-character regions, the segmentation network is trained. The approach then decodes the character

sequence from the pixel-level predictions using a segmentation-based recognition module. The unit of detection is frequently set at the word level in many text detection and identification techniques, where a "word" is defined as a group of letters separated by spaces. However, depending on the language, typeface, and text layout, there may be changes in how words are defined and split, making the borders of certain terms not always clear. Recent text detection and identification techniques have investigated the use of more adaptable and flexible strategies, like character-level recognition and segmentation-based techniques, to get around these difficulties. Instead of depending on the preset word boundaries, these methods try to find and identify specific characters inside a text region. Some studies, such as [26] and [14] propose using surface-level or situational features of the scenic text, such as indentation, curvature, matte finishes, and light reflection, to detect individual characters. But to produce character recognition mape, the study in [27] uses word regions. The TextSpotter [25] is another method that finds text in scenic photographs and recognizes it using a character-level probability map, which is an improvement over WordSup [28] which uses anchor boxes and their sizes to detect text regions. However, scene-based images can have unexpected conditions such as curvature, loops, branches, and embossed text with blurry or shiny backgrounds, which can make character-level detection more challenging. In above mentioned studies, the approaches which have been used by different researchers still lacks in terms of shape and structure of the characters particularly while using Arabic language. Similarly, all these studies for scenic text detection and recognition which have been applied to different languages cause low performance results when applied to the Arabic language. In order to carry out our proposed research study, for feature extraction, we applied transfer learning and reused the VGG-16 pre-trained model (on a sizable dataset). One thousand features are produced each image by the VGG-16 encoder stage. The amount of features is dependent on the number of CNN stages, not the size of the image. These features are trained or used by the decoder stage to provide text/link scores images for the words and characters in the Arabic text lines that were recognized. We are presently interested in producing the character and connectivity scores. The subsequent phase uses these scores to produce word and letter level boundary boxes. The VGG-16 pre-trained model is used to complete the entire process. Therefore, we concentrate on extraction of Arabic text and word boundaries awareness from scene images with different text orientations using enriched VGG-16 based Transfer Learning Auto encoder Network Model [8] followed by Arabic Character Segmentation and Recognition (ACSR) using conventional two-layer Neural Network (NN-2L) Model.

3. Proposed Methodology

The Arabic Word Affinity Region Detection (AWARD) method proposed in this study mainly focuses on detecting and recognizing text embedded in images that have a low background contrast, cursive nature of Arabic characters with different fonts on reflected medium in scenic images, and other related challenges. The primary components of the AWARD technique are image preprocessing, line and word segmentation, bounding box construction, character scaling, and recognition using an enhanced deep neural network model. The input image is enhanced during the preprocessing stage to increase contrast and clarity, and noise reduction techniques may be used to get rid of any extraneous data. The AWARD approach combines preprocessing, segmentation, and deep learning techniques to produce accurate recognition results. It is designed basically aimed at to keep abreast of readers as well as research scholars to get acquainted with dealing overlaid text especially Arabic language to process scene images. The block diagram of the proposed AWARD methodology is presented below see **Fig.**

2 along with brief introduction of the main techniques.

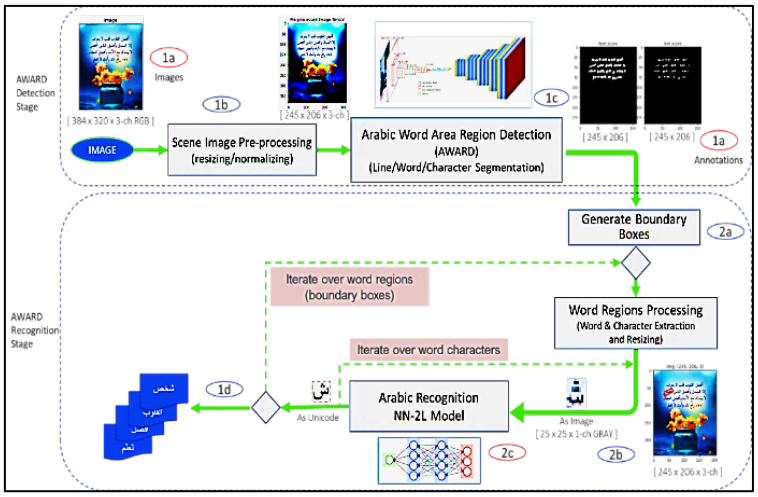

Fig. 2. Block diagram of the proposed methodology design

In the above **Fig. 2**, two sub phases namely detection and recognition are depicted with their respective steps likewise a, b numbered in a sequence. Wherein sub phase "1 detection" has further a, b, c steps representing synthetic dataset generation, scene image processing and score generation respectively. Similarly, the sub phase "2 recognition" shows four steps numbered in a sequence a, b, c, and d to represent segmentation, character cropping and conditioning, Arabic character recognition and Arabic word assembly respectively. Further the proposed methodology design is divided into the following sub-phases and their steps to be followed in sequence.

3.1 Detection Phase

The main component of the AWARD detection stage is a deep convolutional neural network model, which is trained to directly predict the existence of Arabic text instances and the bounding boxes geometries of Arabic characters and words. In order to proceed this phase, at first research gaps has been explored on the basis of previous literature. Then by scrutinizing the past methods on technical grounds, some are excluded in the current phase too. It is pertinent to mention here that the approaches based on segmentation outperformed in terms of accuracy that were somehow skipped in past by others while proceedings detection phase. Previously, accuracy was affected due to the overlapping of the characters in the approached which were supposed to work well over or under segmentation. This presents an x-Ray scan tool that can localize the words and characters on Arabic scene images as depicted in **Fig. 3** below. Further, it consists of three sub-functions namely; Arabic scene images dataset for detection model (synthetic dataset generation), Scene image pre-processing and, Arabic word and character scores generation (CNN Auto-encoder Model).

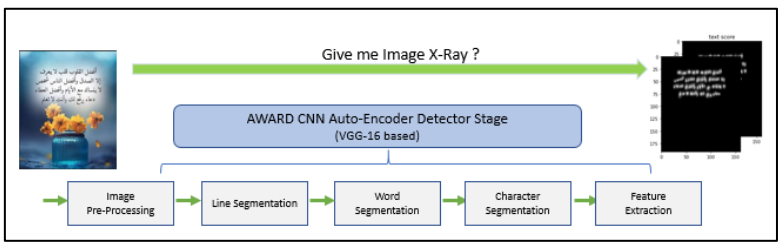

Fig. 3. Concept of affinity box generation from character bounding boxes

3.1.1 Arabic word Detection and Score Generation

The Arabic scene text picture is loaded as an RGB (3-channel) image at the beginning of the AWARD Detection step. Billboards, signs, computer graphics, and even streaming media can all be used to capture images for the RGB representation.The image is converted to RGB if it is grayscale (1 channel) (3-channel) and the standard channels are also eliminated if the image has additional channels, such as transparency. We avoid scaling photos over the threshold n, where n=1280, in order to improve object detection by the VGG-16 encoder (of either height or width), the model loss function, each channel of the RGB scaled image is normalised using mean and variance. For this purpose, two empty canvas to initialize text and affinity/link score maps is created with image dimensions and image is copied to both. The 2 canvases created in the pre-processing step is overlayed on same image structure and passed into the CNN Auto-encoder, which performs image transformation to generate the "link scores" that will be used for calculating word boundaries and "text scores" to calculate character boundaries. The "image transformation" is learned using training step using the generated synthetic dataset.

3.2 Recognition Phase

The main component of the AWARD "Recognition" stage is a using two layers Feed Forward Neural Networks as a classifier, like [29]. In the "Detection Phase," the proposed model is trained by using segmentation either character or word for predicting Arabic alphabets. The process by which Arabic characters are extracted from scene images and are recognized is shown in **Fig. 4** below. Word and character segmentation utilizing affinity/text scores (character/word bounding boxes generation), recognition image pre-processing (character cropping and conditioning), Arabic character recognition, and Arabic word assembly are all included in the list of sub-phases. The procedure begins with the input of scene photos with Arabic language content. Subsequently, the process generates scores that are then input into the word and character segmentation stage, where Arabic words are then formed following recognition.

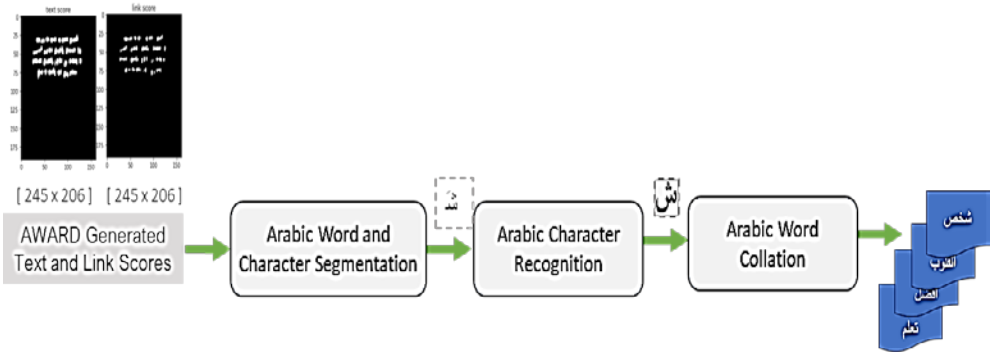

Fig. 4. AWARD "Recognition" process for Arabic character and words

4. Experimentation

The Arabic Character Recognition Model (ACRM) first understands the dataset and then cover Arabic characteristics for the purpose to train the model. As contribution of this research study, a dataset has been created on the basis of visual inspection of results and then quantitative analysis. This dataset comprised on overlaid text found on various scene image of different sizes; and it consists of 5000 cropped images with embedded Arabic text. Similarly, labels are generated from Arabic words' corpus of size 10000 words. The **Algorithm 1** presents an

algorithm used to execute the proposed methodology and subsequently **Table 1**, shows our dataset and the 29 Arabic language characters.

Algorithm 1: Synthetic dataset preparation (D,A,i)

```
INPUT:  Dataset of background Scene Images D; Dataset of Arabic Words A
Output: Annotated Image with Arabic Words
INITIALIZATION:
        D = [5000]          [Images from Google Images]
        A = [10000]         [Arabic Word Corpus]
FOR EACH Image i IN D:
        IF i has Arabic text THEN
        Textloc[i]= max(word_region(image(x,y,h,w))
        ArabicText[i]= affinityScore(TextLoc[i])
        END IF
END FOR
FOR EACH i IN TextLoc:
FeatureImages[images]= Image[i].textLoc[i]transformText(horizontal)
END FOR
FOR EACH fi IN FeatureImage
        IF fi <> RGB THEN
        Convert to RGB
        END IF
        IF fi.resolution< 1280 THEN
        ReImages= fi [ m * max(h, w) ] × fi(h,w,c)        #Resize image to max 1280
        fiheat = fi [Val_x − μ_x / σ²_x]                  #Find Heat Dimension
        END IF
END FOR
#Word Extraction and Segmentation
FOR EACH i IN ReImages:
i = i.convert(binary)
i = i.findConnectedRegions();
WordRegions[i][text] = CroppedWordArea()
END FOR
FOR EACH wr IN WordRegions:
i = 0
FOR txt IN wr:
        Textlabel[i] = Label[wr[txt])
END FOR
Word = Horizontal(texlabel)
AnnotatedImageSet[wr] = word
END FOR
RETURN AnnotatedImageSet;
```

Arabic characters presents many challenges for OCR applications as detailed in [30]. Moreover, it has been found that the Arabic language is semi-cursive in nature which makes it too complicated in the sense of character detection and recognition. Mostly, penmanship style in Arabic language is followed in which symbols are written in flowing manner. That's why it is also known as cursive language but 06 among 28 of its characters are also follow printed blocked letter style which is called as non-cursive [31]. Below are the various characteristics included in the design of the used dataset for AWARD "Recognition" phase model. In the above Algorithm, "Textloc" is an array contains the embedded text locations information of text regions in the image, "i" is an index, "word_region" is a function that takes an image as its input and return a region of the image that contains text. "image" is input image that is being processed while variables "x" is x-coordinate, "y" is y-coordinate," "h" is height "w" represents width of the region of interest in the image. While in the expression "IF fi.resolution< 1280 THEN ReImages= fi [m * max(h, w)] × fi(h,w,c)," image is represented by "fi," "resolution" is resolution of the image, "Reimages" is the output image, "m" is scaling factor, "h" for height "w" for width and "c" is used for number of channels (e.g., RGB) in the input image. We'll use the mapping between character number (No: First Column) to identify

the Arabic character (class) detected. This also maps to either the Unicode representation of the Arabic character.

Table 1. Arabic recognition dataset with characteristics and schema

No	Letter Name	Letter Shape				Dataset Size	Unicode
		Isolated	*Initial*	*Middle*	*End*		
1	Hamza	ء	أ ، ئ	أ ، ؤ ، ئ	ء ، ؤ ، ئ ، أ	4205	U+0674
2	Ba’a	ب	بـ	بـ ، بـ	ـب ، ب	4106	U+0628
3	Ta’a	ت ، ة	تـ	تـ ، تـ	ـت ، ت ، ـة	4273	U+062a
4	Theh	ث	ثـ	ثـ ، ثـ	ـث ، ث	4249	U+062b
5	Jeem	ج	جـ	جـ ، جـ	ـج ، ج	4453	U+062c
6	Ha’a	ح	حـ	حـ ، حـ	ـح ، ح	4459	U+062d
7	Kha’a	خ	خـ	خـ ، خـ	ـخ ، خ	7630	U+062e
8	Dal	د	د	ـد ، د	ـد ، د	5138	U+062f
9	Thal	ذ	ذ	ـذ ، ذ	ـذ ، ذ	4366	U+0630
10	Ra’a	ر	ر	ـر ، ر	ـر ، ر	5411	U+0631
11	Zai	ز	ز	ـز ، ز	ـز ، ز	4473	U+0632
12	Seen	س	سـ	سـ ، سـ	ـس ، س	4305	U+0633
13	Sheen	ش	شـ	شـ ، شـ	ـش ، ش	4343	U+0634
14	Sad	ص	صـ	صـ ، صـ	ـص ، ص	4429	U+0635
15	Dad	ض	ضـ	ضـ ، ضـ	ـض ، ض	4450	U+0636
16	Tah	ط	طـ	طـ ، طـ	ـط ، ط	4259	U+0637
17	Dtha	ظ	ظـ	ظـ ، ظـ	ـظ ، ظ	4397	U+0638
18	Ain	ع	عـ	عـ ، عـ	ـع ، ع	4322	U+0639
19	Ghain	غ	غـ	غـ ، غـ	ـغ ، غ	4381	U+063a
20	Fa’a	ف	فـ	فـ ، فـ	ـف ، ف	4438	U+0641
21	Ghaf	ق	قـ	قـ ، قـ	ـق ، ق	4368	U+0642
22	Kaf	ك	كـ	كـ ، كـ	ـك ، ك	4387	U+0643
23	Lam	ل	لـ	لـ ، لـ	ـل ، ل	4392	U+0644
24	Meem	م	مـ	مـ ، مـ	ـم ، م	4426	U+0645
25	Noun	ن	نـ	نـ ، نـ	ـن ، ن	4588	U+0646
26	Ha’a	ه	هـ	هـ ، هـ	ـه ، ه	4555	U+0647
27	Waw	و	و	ـو ، و	ـو ، و	4317	U+0648
28	Alef	ا ، آ	ا ، آ	ا ، آ ، ـا	ا ، آ ، ـا	4303	U+0627
29	Ya’a	ي	يـ	يـ ، يـ	ـي ، ي	4392	U+064a

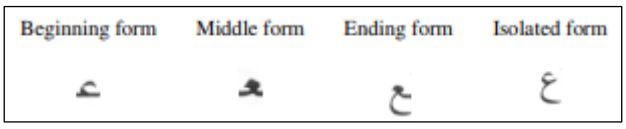

Fig. 5. (a) how the same Arabic character with 4 different shapes

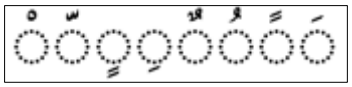

Fig. 5. (b) Same Arabic letters pronunciation indicators

5. Evaluation Criteria and Performance

To evaluate the performance of the AWARD "Recognition" model we used the SciKit-Learn python library to implement the 2-hidden layers MLP model and using the 70% to 30% training to validation split. We used confusion matrix reporting the observations true/false positives and negatives (TP, FP, TN and FN) and using the resulting classification metrics.

The AWARD "recognition" model performance is reported in last row of the **Table 2**. For each Arabic character/class, and **Table 2** presents an overall summary below separately:

Table 2. Performance summary of AWARD recognition model

Character/Class	Precision	Recall	f1-score	Support
ا	93.9	94.7	94.3	611
ب	93.2	90.9	92	584
ت	94.4	93.1	93.7	637
ث	93.9	94.8	94.4	606
ج	94.3	94.7	94.5	565
ح	94.2	94.4	94.3	619
خ	94.1	94.4	94.2	630
د	94.8	94.4	94.6	618
ذ	94.4	94.7	94.5	599
ر	94.5	94.4	94.4	590
ز	94.5	94.3	94.4	580
س	94.7	95	94.8	621
ش	94.5	94.8	94.7	596
ص	94.8	94.2	94.5	603
ض	94.8	94.5	94.7	598
ط	94.4	94.7	94.5	620
ظ	94.8	94.5	94.7	605
ع	93.5	94.7	94.1	582
غ	94.2	94.4	94.3	615
ف	94.2	93.6	93.9	601
ق	94.4	93.9	94.1	591
ك	95	95	95	612
ل	94.2	94	94.1	588
م	94.8	94.7	94.8	585
ن	94.7	94	94.3	586
ه	94.2	93.1	93.6	595
و	94.2	94	94.1	597
لا	94.7	94.7	94.7	578
ي	90.5	94.4	92.4	588
Average Measure	**94.2**	**94.2**	**94.2**	**-**

The tradeoff between the true positive rate and the false positive rate can be visualized in the Receiver Operating Characteristic (ROC) curve of the model. **Fig. 6(a&b)** shows the AUC (Area under Curve) analysis using "ovo" model "one.vs.one" per Arabic character which shows the high classification match performance:

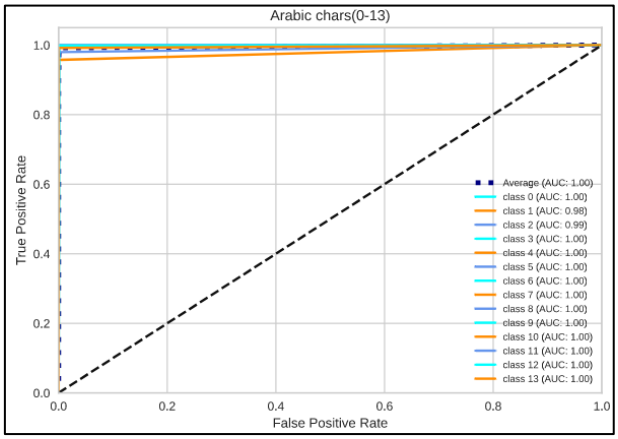

Fig. 6. (a) AWARD "Recognizer" model curve

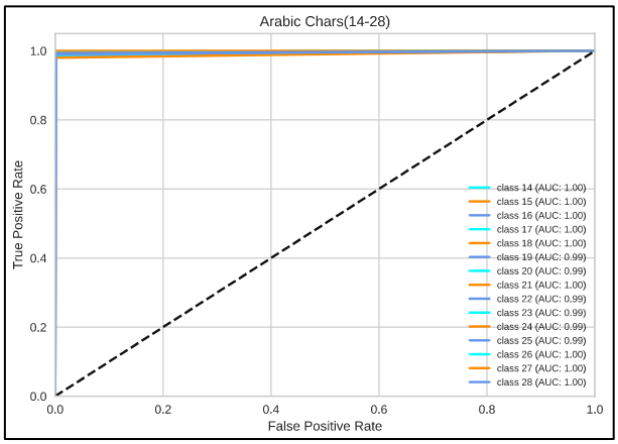

Fig. 6. (b) AWARD model ROC curve for "one.vs.one"

Linear SVM, I-Layer NN, Gaussian Naive Bayes, and 2-Layer NN are tested on the data with different variations and 25% of the entire data set. In this steps, the achieved accuracy is tested through several trials to get better understanding about the performance of the proposed method. Further, the accuracy of the aforementioned networks is shown in **Fig. 7** below wherein an activation function is used with a hidden 200 x 100 layer in size.

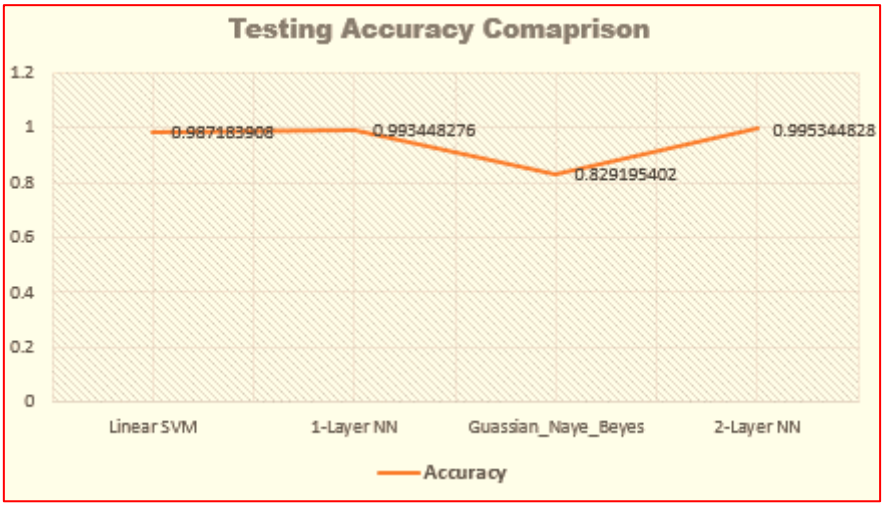

Fig. 7. Presents the accuracy of various networks tested on 25% of the same dataset

5. Conclusion

This study explicitly revealed the pre-processing phase of the images as to tune the dataset to be used on the proposed Arabic Word Awareness Region Detection (AWARD) platform. In pre-processing phase the various steps were applied as the requirements of the reused VGG-16 pre trained model. In this study, we mainly concentrated on the creation of synthetic data set and presents the results of images based on visual analysis and quantification. After experimentation, AWARD detects words and characters in the scene Arabic text images and produces the word region score and the character affinity scores which together covers the various shapes of Arabic characters. The bounding boxes at character level will prepare this framework for its next step of segmentation that Arabic characters in the scene text are properly recognized by a two layer neural network. In order to achieve the desired objectives we have used the activation function is le-05 with hidden layer size (200,100) and with maximum iteration 1000. To conclude the discussion and to give clear the view point to the readers about the Arabic words recognition through scenic images, the following points are considered and significant results are obtained to be utilized and improved in future: creation of synthetic dataset, preprocess scene input text images to remove or reduce noise and emphasize Arabic words text, detect the Arabic word region and character locations on text scene images using convolutional auto-encoder based on pre-trained VGG-16 based model, apply text characters scores and linkage scores generated by VGG-16 model to extract boundary boxes on Arabic words and characters and recognize the Arabic characters and words in scene text images using improved traditional NN learning model. Resultantly, our proposed approach achieved significant results with 91.8% word segmentation accuracy and 94.2% character recognition accuracy. Potential image processing and feature extraction researchers will find it useful to analyze scene mages in any language while also considering text images to improve or reduce noise by employing neural networks to expand the capability of a VGG-16 based model.

References

[1] Khorsheed, M. S., "Off-line Arabic character recognition-a review," *Pattern analysis & applications*, vol. 5, no.1, pp. 31-45, 2002 Article (CrossRef Link)

[2] Alkhateeb, J. H., J. Ren, S. S. Ipson, and J. Jiang, "Knowledge-based baseline detection and optimal thresholding for words segmentation in efficient pre-processing of handwritten Arabic text," in *Proc. of International Conference on Information Technology: New Generations*, Las Vegas, NV, 2008. Article (CrossRef Link)

[3] Khorsheed, M. S. and W. F. Clocksin, "Structural Features of Cursive Arabic Script," in *Proc. of BMVC*, pp. 1-10, 1999. Article (CrossRef Link)

[4] Lorigo, L. M. and V. Govindaraju, "Offline Arabic handwriting recognition: a survey," *IEEE transactions on pattern analysis and machine intelligence*, vol. 28, no.5, pp. 712-724, 2006. Article (CrossRef Link)

[5] Zeki, A. M., "The segmentation problem in arabic character recognition the state of the art," in *Proc. of ICICT'05*, pp. 11-26, 2005. Article (CrossRef Link)

[6] Baek, Y., B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in *Proc. of CVPR'19*, pp. 9357-9366, 2019. Article (CrossRef Link)

[7] Nazif, A., "A system for the recognition of the printed Arabic characters," Master's Thesis (2nd Edition) Faculty of Engineering, Cairo University, 1975.

[8] Guan, Q., Y. Wang, B. Ping, D. Li, J. Du, Y. Qin, H. Lu, X. Wan, and J. Xiang, "Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: a pilot study," *Journal of Cancer*, vol. 10, no. 20, pp. 4876-4882, 2019. Article (CrossRef Link)

[9] Huang, H. and F. Da, "Sparse representation-based classification algorithm for optical Tibetan character recognition," *Optik*, vol. 125, no.3, pp. 1034-1037, 2014. Article (CrossRef Link)

[10] Khorsheed, M. S., "Automatic recognition of words in Arabic manuscripts," University of Cambridge, Computer Laboratory, 2000.

[11] Alshebeili, S. A., A. a.-F. Nabawi, and S. A. Mahmoud, "Arabic character recognition using 1-D slices of the character spectrum," *Signal Processing*, vol. 56, no.1, pp. 59-75, 1997. Article (CrossRef Link)

[12] Al-Shatnawi, A. M. and K. Omar, "The Thinning Problem in Arabic Text Recognition-A Comprehensive Review," *International Journal of Computer Applications*, vol. 103, no.3, pp. 35-42, 2014. Article (CrossRef Link)

[13] Amin, A., "Recognition of hand-printed characters based on structural description and inductive logic programming," *Pattern recognition letters*, vol. 24, no. 16, pp. 3187-3196, 2003. Article (CrossRef Link)

[14] Safabakhsh, R. and P. Adibi, "Nastaaligh handwritten word recognition using a continuous-density variable-duration HMM," *Arabian Journal for Science and Engineering*, vol. 30, no.1, pp. 95-120, 2005.

[15] Farooq, F., V. Govindaraju, and M. Perrone, "Pre-processing methods for handwritten Arabic documents," in *Proc. of ICDAR'05*, pp. 267-271, 2005. Article (CrossRef Link)

[16] Alginahi, Y. M., "A survey on Arabic character segmentation," *International Journal on Document Analysis and Recognition (IJDAR)*, vol. 16, no.2, pp. 105-126, 2013. Article (CrossRef Link)

[17] Khurshid, K., C. Faure, and N. Vincent, "Feature-based Word Spotting in Ancient Printed Documents," in *Proc. of PRIS*, pp. 193-198, 2008.

[18] Märgner, V. and H. E. Abed, "Databases and competitions: strategies to improve Arabic recognition systems," in *Proc. of Summit on Arabic and Chinese Handwriting Recognition*, pp. 82-103, 2006. Article (CrossRef Link)

[19] Wirotius, M., A. Seropian, and N. Vincent, "Writer identification from gray level distribution," in *Proc. of ICDAR'03*, pp. 1168-1168, 2003.

[20] Broumandnia, A., J. Shanbehzadeh, and M. Nourani, "Handwritten farsi/arabic word recognition," in *Proc. of IEEE-ICCSAs*, pp. 767-771, 2007. Article (CrossRef Link)

[21] Khorsheed, M. S., "Hmm-based system for recognizing words in historical arabic manuscript," *International Journal of Robotics and Automation*, vol. 22, no.4, pp. 294-303, 2007. Article (CrossRef Link)
[22] Seeger, M. and C. Dance, "Binarising camera images for OCR," in *Proc. of ICDAR'21*, pp. 54-58, 2001.
[23] Al-Rashaideh, H., "Preprocessing phase for Arabic word handwritten recognition," *Information Process (Russian)*, vol. 6, no.1, 2006.
[24] Paquet, T. and Y. Lecourtier, "Automatic reading of the literal amount of bank checks," *Machine Vision and Applications*, vol. 6, no. 2, pp. 151-162, 1993. Article (CrossRef Link)
[25] Sauvola, J., T. Seppanen, S. Haapakoski, and M. Pietikainen, "Adaptive document binarization," in *Proc. of ICDAR'97*, pp. 147-152, 1997.
[26] Tellache, M., M. Sid-Ahmed, and B. Abaza, "Thinning algorithms for Arabic OCR," in *Proc. of IEEE Pacific Rim Conference on Communications Computers and Signal Processing*, pp. 248-251, 1993.
[27] Sari, T., L. Souici, and M. Sellami, "Off-line handwritten Arabic character segmentation algorithm: ACSA," in *Proc. of Eighth International Workshop on Frontiers in Handwriting Recognition*, pp. 452-457, 2002.
[28] Kimura, F., "Context directed handwritten word recognition for postal service applications," in *Proc. of ATC'92*, Walsh. DC, pp. 199-213, 1992.
[29] He, P., W. Huang, T. He, Q. Zhu, Y. Qiao, and X. Li, "Single shot text detector with regional attention," in *Proc. of ICCV'17*, pp. 3066-3074, 2017. Article (CrossRef Link)
[30] Alkhateeb, J. H., J. Ren, S. S. Ipson, and J. Jiang, "Knowledge-based baseline detection and optimal thresholding for words segmentation in efficient pre-processing of handwritten Arabic text," in *Proc. of Fifth International Conference on Information Technology: New Generations (itng 2008)*, pp. 1158-1159, 2008. Article (CrossRef Link)
[31] Al-Ohali, Y., M. Cheriet, and C. Suen, "Databases for recognition of handwritten Arabic cheques," *Pattern Recognition*, vol. 36, no.1, pp. 111-121, 2003. Article(CrossRef Link)

Ayed AlRadaideh is a PhD student in the UTM ViCubeLab Research Group (VICUBE) at Faculty of Computing, Universiti Teknologi Malaysia. He received his BSc in Computer Science from Yarmouk University, MSc Computer Science from Yarmouk University, Jordan. His research includes Image Processing, Machine Learning, Deep Learning and Computer vision.

Mohd Shafry Mohd Rahim is a Professor of Image Processing at School of Computing, Faculty of Engineering, University Technology Malaysia, Skudai, Johor, Malaysia. Presently, he has appointed as Chair of Institute for Life Ready Graduate, University Technology Malaysia (UTM iLeaGue) from Jun 2020 and also as a Research Fellow of Media and Game Innovation Centre of Excellence (MaGICX), Institute of Human-Centred Engineering (iHuMEN), University Technology Malaysia. Besides, he is member of the Board of Governance (BOG), SPACE College since 2014. He received his Diploma in Computer Science (1997), B.Sc.of Computer Science majoring in Computer Graphics (1999), and MSc. Of Computer Science (2004) from the University Technology Malaysia (UTM), Malaysia and his PhD of Spatial Modelling (2008) from University Putra Malaysia (UPM), Malaysia.

Wad Ghaban is an assistant professor in the applied college at the University of Tabuk, Saudi Arabia. Wad received her BSc in computer science from King Abdul-Aziz University in Jeddah with honors degree. Then, she received her MSc. in advanced computer science with distinction in University of Birmingham by 2015. Later, she got her PhD from University of Birmingham by 2020. During her study, Wad worked on several projects related to Human computer interaction, survival analysis, online learning, and Natural language processing and sentiment analysis. Was also published a number of papers that are published and presented in several international conferences and indexed journals. Her research interests are human computer interaction, machine learning, sentiment analysis and data analysis.

Shahid Kamal is a self-employed computer science researcher in D.I.Khan Pakistan since 2020. He has received his PhD degree in computer science with major in Software Engineering /information retrieval from Faculty of Computing, Universiti Teknologi Malaysia in 2016. He is the member of Software Engineering Research Group (SERG). He has served Gomal University as faculty member in capacity of Lecturer and Assistant Professor as well for above 15 years. His research includes information systems, data mining, web search and its related issues and information integration. Besides, he is active team member in research projects likewise Government big data Ecosystems, AI Applications & Future. Dr. Shahid Kamal has 29+ publications in journals, Proceedings, and Book Chapters.

Naveed Abbas is an Assistant Professor of Image Processing at Department of Computer Science, Faculty of Engineering and Technologies, Islamia College University Peshawar, KPK, Pakistan. Presently, he has been appointed as Chair of the Computer Science Department since August, 2022. He received his Diploma in Computer Science and I.T (2001), Bachelor of Computer Science (04years) majoring in Digital Image Processing (2006), MS Computer Science (2012) from the Islamia College University Peshawar, Pakistan and his PhD Computer Science specialization in Medical Imaging and Machine Learning (2016) from University Technology Malaysia (UTM), JB, Malaysia.