Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

Survey of Multilingual Script Identification Techniques on Wild Images

Kiran Perveen[1], Rukhsana Perveen[2], Dr. Awais Yasin[3]
[1]Department of Computer Science, COMSATS Institute of Information Technology (CIIT), Islamabad.
[2]Department of Electrical Engineering, National University of Computer and Emerging Sciences (FAST), Islamabad.
[3]Department of Computer Engineering, National University of Technology (NUTECH), Islamabad-Pakistan.
kiranch3977@gmail.com, rrukhsana9@gmail.com, awaisfrombit@gmail.com

DOI: 10.5281/zenodo.6547188

ABSTRACT

Multilingual Script Identification on natural images has recently increase research attention and this is very challenging task. This paper presents a review of latest techniques for the multilingual scripts. The system can choose the appropriate Optical Character Recognition (OCR) engine to recognize a script based here on script identity of a retrieved line of text or word. A number of approaches for identifying different characters, including such Japanese, Chinese, Arabic, Korean and Indian, have been developed. scripts are used in written on natural scenes captured by a voyager from cameras or text recognitions system. Here we also present the difficulties that come with script identification, methods used for features extraction and also the classifiers used for identification. We provided a comprehensive description and evaluation of previous and state-of-the-art script identification approaches. It should be emphasized that researchers in the area of multilingual script recognition is still in its early stages, and additional analysis is needed.

Keywords: Latest script, wild images, style, script techniques, challenges images.

Cite as: Kiran Perveen, Rukhsana Perveen, Awais Yasin. (2022). Survey of Multilingual Script Identification Techniques on Wild Images. LC International Journal of STEM (ISSN: 2708-7123), 3(1), 1–14. https://doi.org/10.5281/zenodo.6547188

INTRODUCTION

Multilingual Script Identification is an important area of research and the subfield of Pattern Analysis and Recognition. Scripts is just a graphic representation of a written tradition that is used to represent textual cultures. The script would be used by just one speech or by a variety of languages, with minor differences from one to another. For example European languages like English, German, French and some other languages use different variants of Latin alphabets and so on. Therefore there is an increased demand for Multilingual Script Identification because a lot of data on natural images, sign boards, besides the roads on banners, street walls etc. They were already in written text and required to be fed into a computer system for additional processing by an explorer using a digital camera. As a result, in today's multilingual world optical character recognition (OCR) systems must first be capable of identifying the script in which they are written, followed by text recognition, language recognition, and other processes. Script recognition on paper and electronic text documents has been a highly focused problem for the last two decades and continues to be so today D. Ghosh (2010), K. Singh(2015). Multilingual script identification in natural scene is relatively new aspect of script identification and nowadays need more accuracy in this field. Script recognition in scene photographs is difficult due to the following factors:

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

(i) The text in a scene image frequently begins with a single phrase or a sequence of characters. (ii) Stylish and variety of fonts in Scene text images to Attract more visitors and make it difficult to recognize the writing style. (iii) Images have different aspect ratios (iv) Script come with highly complex background images often contain predominantly text. As a result, text translation and managing with false alarms are two other issues.

LITERATURE REVIEW

Need of Script Identification

1) For the creation of automated comprehension structures built for the multilingual society, script recognition from a natural scene images is a must.
2) It is essential to recognize the scripts before using appropriate text extract or classification, and Optical Character Reader (OCR) methods to recover the text from photos or videos from national and foreign sources Obaidullah et al. (2018c).
3) Previously known script recognition techniques relied on recognizing script in machine-printed scanned documents or handwritten images with a simple or plain background. As a result, an attempt to expand the work to more demanding cases is required due to a lack of time.
4) The Text Separation and Binarization (TSB) method is affected by language-dependent factors such as stroke statistics, stroke density, aspect ratio, and font size. As a result, prior understanding of the script is needed.
5) Because some languages, such as English, Greek, and Russian, may share a common character set, it is necessary to capture local information for discrimination, as global knowledge is not always sufficient Bhunia et al. (2019).
6) The majority of the documented works in the field are for binary picture recognition, which has an impact on the process because converting an image representation to digital may lose important parts of the text.

In Optical Character Recognition (OCR) system, The very first step in detecting the dialect where a text picture or content is created is to identify the script. This phase is indeed required for the postprocessing of the multilingual text image or document, such as routing, indexing or inside the text, translation is preceded by speech perception. Inside a multi-script context, script detection can aid in the detection of text areas in video indexing and retrieval, as well as material categorization in library resources.

Automatic script identification from documents or from images has many important applications such as getting information from unknown script, sorting information, Interpretation of foreign texts and categorization of a vast database of these images When executing an individual process, script recognition is critical to the success language OCR. Thus, script identification has a great impact in OCR system for understanding multilingual script, on natural scenes, images and documents for creating a digital library of literature written in several scripts In multiple languages scenarios, script identifying in image features necessitates the required pre - processing of a visual feature comprehension system M. Kundu (2010), D. Karatzas (2013) and J. Makhoul (1999), that is Scenery comprehension R. Socher(2009), smartphone navigating T. Cour(2008), film subtitle identification V. Alabau (2014), and translation software A.H. Toselli(2010) are only a few examples.

Previous character recognition research has primarily focused on texts T.N. Tan (1998), A. Busch (2005) and G.D. Joshi (2007) and movies Z. Ding (2011) and D. Zhao (2012). In Ghosh et al(2010) made a detailed and precise survey on script identification at the level of the illustrate that the proposed, text-line, sentence or content, or term. Classification of text images can do by their To extract some form of holistic aspect characteristics from the input image, feature extraction is used. Tan et al(1998) provides

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

texture properties that are rotational independent method for feature extraction method. to identifying document scripts. We give a detailed analysis of multilingual script processing algorithms created primarily for the detection of several important world scripts, such as Latin, Korean, Chinese, Japanese, Hebrew, Arabic, Cyrillic, and the Almost every family of Indian characters, in this work. In Section 2, we give a precise description of latest Multilingual Script Identification Methodologies that represents their main distinguish values. Comparative analysis mentioned in section 3 and in section 4 we conclude this review of multilingual script identification.

METHODOLOGY

Multilingual Script Identification Methodologies

Convolutional Neural Network (CNN) variation for character recognition (MSPN), B. Shi et al. (2015) proposed the Multi-stage Spatially-sensitive Pooling Network. This network focuses on a deep learning approach for multilingual realistic scene scripts detection at the word or line level, as well as input photos with various dimensions. As in a traditional CNN structure, CNN Only input with such a set of measurements can be accepted by layers. They suggest the spatially-sensitive pooling (SSP) layer to solve this challenge, that retains useful hierarchy information for script recognition while also dealing with various input sizes. MSPN scheme represent features at different levels. In figure 1, the outputs the three convolutions are combined into a long vector that is fed to later completely levels. This system advantages from the rich and discriminatory properties of the lengthy vector because it contains pooled characteristics for script identification. They also evaluate their MSPN and previous baseline methods on dataset SIW-10 B. Shi et al. (2015) achieves 94.4% performance than traditional approaches including traditional methods Y. LeCun (1998) and Locality-constraint linear coding [20].

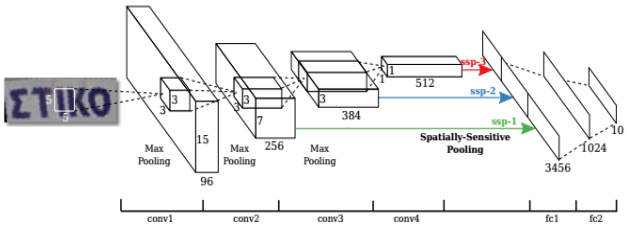

Figure 1: MSPN architecture illustrated. These multiple arrows indicate the spatially sensitive pooling layers. The vector length of probability SoftMax output value.

This technique performed identification in cropped word images, gives a new and promising direction where is to look at methods to detect the linguistic type of words using healthy long images instead of cropped images. A. Ul-Hassan et al. (2015) The sequential properties of Long Short-Term Memory (LSTM) systems are used to propose a novel way for numerous character recognition. The spatiality of their work is wwhenever two or even more languages will be in the same message, to distinguish between them. This technique need no preprocessing technique nor feature extraction step. For a given script, trained the LSTM network for all alphabets to master a key target category. They used open source OCRopus(2021) for their experiments. In training phase, without feature extraction and ground-truth information, they provide ground-truth is a succession of category. The target protein for a specific text lines graphic is then changed ground-truth knowledge. The Network model attempts to learn the forms of specific characters after first locating the gaps in the text as shown in figure 2.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

Figure 2: The LSTM trained model output from the for identification of script.

They used self-made dataset that contain English-Greek text by using OCRopus module to generated 90,000 images of different types of fonts, 1D-LSTM network is same as T. M. Breuel (2013), a sliding window of 1-pixel width traverse the images to convert them into text-line sequence that fed to LSTM network for training session. This network consist of input layer one, hidden layer having one output layer and 50 cells. With the one LSTM training stage, a moving window of 481 passes across the source images, so each frame is passed to the LSTM network during learning. It made advantage of bidirectional structures, with multiple convolutional layers, one scanning the inputs from right to left and another from left to right, and a single output layer connecting all layer upon layer. Applying the1D-LSTM model and performance of an accuracy is 98.19%, LSTM networks' tremendous learning capabilities are demonstrated. The conclusion is that LSTM networks are capable of learning several various forms for same class also useful for large-scale classification problems where have large variations.

Apart from traditional methods for script identification on document images, or paragraph images, A. K. Singh et al. (2015) proposes a simple and an efficient method for the Without even any extra adjustment, script and country detection is possible of input word or line images. In precise way, the training the RNNs to consistently differentiate the script or speech by converting individual phrase or line pictures into a collection of extracted features. It had used profile features P. Krishnan (2014) and R. Mamatha (2003) from an unsegmented word image, it comprises of two LSTM networks and is based on a Backpropagation Long Short-Term Memory (BLSTM) system. One system receives data from the start to the finish, whereas the other receives data from the finish to the start. Final outputs will predict by using of the single output of LSTM approaches. For final identification step, the Just at output nodes of the RNN system, Connectionist Temporal Classification (CTC) [27] is used. They utilized 960K phrases and 240K phrases from all of the scripts to train the System for word level character recognition. A different network (with the same architecture) is trained with 120K lines at the operation level, followed by verification using 60K lines from across all scripts. On the provided data, they carried out experiments at the lines and phrase level 12-indic script has average accuracy 95.55%.

A. Nicolaou et al. (2006) proposed a Hand-crafted attribute and powerful Multi-Layer Perceptron (MLP) classification techniques are used to identify scripts. Stated that that Script identification (SI) means focusing Language identification (LI) aims to detect certain supplementary signs, such as punctuation marks, as well as a fundamental learning algorithm. Targeted data belong to video-text and Detection of handwriting text using visual style. Also stated that, the Convolutional neural the disadvantage of using a network technique is that it requires a lot of computer power and a lot of annotations the disadvantage of using a network technique is that it requires a lot of computer power and a lot of annotations. Architecture of their research contains three main parts, (i) Preprocessing step in which input text or word image convert into single channel that is light background and dark foreground. (ii) Local Binary Pattern Sparse Radial Sampling For extracting features, the (SRS)-LBP form of LBP histogram features A. Nicolaou (2015) is used.

The SRS-LBP was obtained for three different regions inside the image data: the various hierarchies, the middle half, as well as the bottom portion of the image and (iii) for classification, as a decoder of the feature extraction to a specified and restricted collection of language, a deep Multi-Layer Perceptron (MLP) was being used. The system is made up of three convolution layers as well as a number of neurons.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

Proposed method used database was CVSI Video Script Identification N. Sharma (2015), it contains 10 different languages that were used Arabic, Bengali, Hindi, Kannada, Oriya, English, Gujrati, Punjabi, Tamil, and Telugu are the varieties spoken in India. The collection is made up of clipped photos with only one word on them. The success is found that the first level of the deeper MLP achieves 98.2%, while layer 2 achieves 97.3 percent, and the activation function achieves 97.9%. Using the same methodology performs on wild scene text detection, using database SIW-10, and SIW-13. The MSPN approach developed in B. Shi (2015) achieves 94.6 percent state-of-the-art efficiency on the given dataset. There are two different datasets which are available publicly, contains cropped word images in ten renowned languages that are SIW-13 B.S hi (2016) adds Cambodian, Kannada, and Mongolian to the languages of SIW, as well as Arabic, Chinese, English, Korean, Russian, Greek, Hebrew, Japanese, Thai, and Tibetan. SIW-10 is divided into a train-set of 8,045 samples and an experiment of 5,000 samples, whereas SIW-13 is divided into 9,791 and 6,500 photos.

A. k. Singh et al. (2016) presented a system that automatically identifying the script Motivated by the development of multi characteristics that are aggregated from intensively calculated feature points in scene photos. This work based on the following challenges; (i) Wild images mostly identified in single word and groups of words, so these images have lack of context information about the text. (ii) Natural scene or wild the use of beautiful fonts in scene text graphics is common to captivate readers, but these fonts do not simply generalize to test results. (iii) Scene text frequently includes a highly sophisticated actual scene context. This method got influence from mid-level features B. Fernando (2014), M. Juneja (2013) and Y. Boureau (2010), where they We calculate the feature points on the specified wild text document, then pool that spatial information to represent the image's broader context. Learn a classifier to recognize the scripts of a training images by representing each initiation and progression with a bag of such mid-level attributes.

The benefits of this work are twofold: first, it is resistant to fluctuations and noise that are frequent in scene text. Second, the method is simple to develop, train, and use numerically. For experiments there is dataset also introduced there are 500 words in this work, titled Indian Language Scene Text (ILST) scene images with more than 3000 words. The text images belong to Telugu, Kannada, Hindi, Tamil, Malayalam and English are the six most regularly used languages in Asia. To be architecture of this work is as follow, firstly local features and densely compute values of given text image by computing the SIFT descriptions at normal grid cells with M image separation. And then pool these extracted features into A rendering of the manuscript at an immediate post. They learn a classification to detect the scripts of a training images by representing each training image with a bag of such post attributes. Using crop words of ILST dataset, achieves scripts Detection performance of 88.67% is much superior than approaches utilized in the image data script recognition area, such as P. B. Pati (2008) and T. Ojala (2008). Experiments on CVSI dataset, proposed method achieve in all 10 scripts, the text responded was 96.70 percent, and there were only 2 techniques in the contest, C-DAC or CUK. L. Gomez et al. (2016) proposed a multi-stage approach for multilingual script identification in natural images by means of combining Naive-Bayes Closest Neighbour classification and neural layers A unique blend of multilayer features A. Coates (2011) and the Naive-Bayes Nearest Neighbour (NBNN) classifier O. Boiman (2008) inspired the proposed technique. The service's goal is to capture powerful central features and then use them in a classifier to maintain the racially discriminatory strength of little distinct images.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

That's why procedure of script identification has influence of fine-grained frame-work B. Yao (2012) and J. Krause (2014). Pre-processing stage contains the resizing input image with aspect ratios and then extracts dense features by using sliding window method. For feature representation, fed extracted features into Naive-Bayes Nearest Neighbour classification and single layer Convolutional Networks (NBNN) is used for classification. The main reason for using it is that, without any intermediary description encoding, it calculates direct image-to-class (I2C) similarity. For Script identification and text recognition, authors introduced the MLe2e multi-script datasets obtained We made each patch 32 x 32 in the very same way as we did in Huang F (2019). As an example, from different datasets. Dataset contains 711 scene images belong to four individual scripts, Chinese, Latin, Hangul, and Kannada. 450 images utilized for training phase, and 261 images were utilized for testing. There is another dataset were used for experiments is N. Sharma (2015). Identification result of CVSI dataset is 97.91 %, and for MLe2e dataset, the result is 91.12%. ResNet-20 Huang F (2019) introduced a Classifier that is used Local and Global CCN features for the aim of understanding automatic text framework which is developed for multi- language environments under the machine learning methods. Pictorial presentations of this designed see in Figure 3. Local CNN uses patches as inputs. Therefore, From the photos, we first select areas of the very same area.

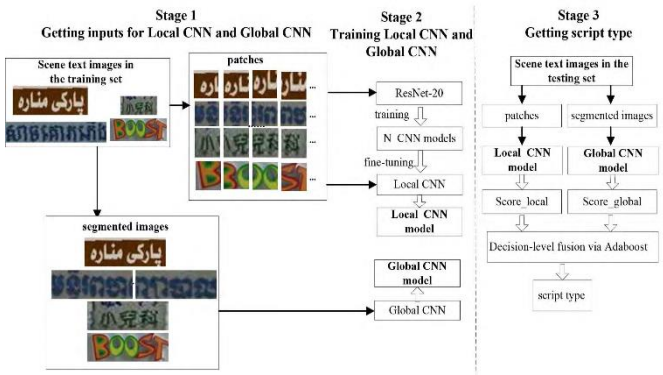

Figure 3: The Pictorial representation of proposed method [43].

We firstly transform it to a grayscale and reduce the height to 40 pixels while maintaining the original screen resolution. Then, out from we definitely extract regions with a size of 32 in the vertical and lateral dimensions of the reduced image. The extracting stage is 8 squares in size. The program carries out the entire patch extraction process automatically without human intervention.

With conventional the input photographs must all be same size, according to CNN. However, because text picture aspect percentages vary so much, scaling the images to the same size causes overall distortion and reduces recognition accuracy. Global CNN is thought to be using split image as an input at this point. Whenever we examine at text lines in photos of visual features, we can see that when the breadth is three times the size, the line's overall qualities are mostly preserved. As a result, we made each segmentation map 40 x 120 pixels in size. We begin by converting an image into different, then increasing the width to 40 pixels while maintaining the screen resolution of the source image If the article's width is much less than 240 pixels after scaling, it is lowered to 120 characters. If the image's width is larger than or equal to 240 pixels, we split it into many 40 x 120-pixel drawings files.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

Algorithm 1 Huang F et al. (2019) presents the full segmentation method.

Algorithm 1 Image Segmentation Process

Input: images of scenario text
Output: image segments
1: **for** each i ∈ [1, *I*] **do**
2: Resizing images I'm going to stick with a height of 40 pixels.
3: D*i* = resized image width
4: if D*i* < 240 then
5: *segImg*(i, 1) = resizing image to [40.120];
6: *splitNumi* = 1;
7: **else**
8: splitNum*i* = D*i*/120
9: *iFirstPos=0;*
10: **for** each *j* ∈ [1, *splitNumi* − 1] do
11: *segImg(i, j)* = imcrop(*iFirstPos*,0);
12: *iFirstPos* = *1* + 120;
13: **end for**
14: the last sub image = imcrop(*iFirstPos*, 0);
15: *segImg*(*i*,*splitNumi*) = resizing image to [40,120];
16: **end if**
17 **end for**
18: **return** *segmentImg(i, j)I =1toI,j=*1*to SplitNumi*

Somiya et.al. (2018) proposed convolutional neural networks (CNNs) with multilevel 2D discrete Haar wavelet transform for feature extraction of images, scaled on variety of different sizes. In this paper, the Bangla, Kannada Malayala, Devanagari, Gujarati, Gurumukhi, Oriya, Roman, Tamil, Telugu, and Urdu are among the 11 written Indian scripts identified using the deep learning method. The author's major motivation was to avoid manufactured identifying elements. CNN based model learns features from images that can contribute to identification/classification of model. She uses twofold: first, he uses for three distinct scales of source images, he employs two and three layered CNNs, and for two different scales of reconstructed image, he employs the identical similar SNNs. CNN architecture of transformed image includes definitions and parameters shown in figure 4. Above figure 4, a CNN's multilayered structure is made up of three basic layers:

1) Convolutional layer (CL)
2) Pooling layer (PL)
3) Fully connected layer (FCL)

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

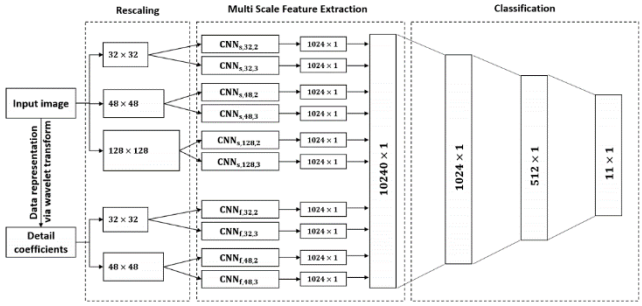

Figure 4: Different mods are shown in this schematic diagram of written Languages character recognition Somiya et.al. (2018).

CLs are made up of a number of kernels that create variables and aid in the combination process. As just an output, every kernel in CL produces an input vector. PLs don't have any characteristics, but their primary goal is to avoid data redundancy (while maintaining their meaning). For their respective PLs, all CNNs in our method have a maximum max pooling. FCL, in which MLP has been incorporated, is employed in addition to these two separates related to surface. Proposed work shown in Fig.4 provide Handwritten indict script design schematic representation in its whole. Someya et al. use 10 different CNNs to extract the feature from a variety of reconstructions of the photo under investigation, referring to each one as CNN d, x, y. In each CNN d, x, y, d and x each refers to the website domain representations and the article's dimension, whereas y means the number of convolution and pooling levels within this CNN. The document known can be represented as d = {s, f} where s denotes the spatial domain and f is the frequencies. In this scenario, one or the other is taken into account. Some wavelengths are muted while applying the Haar Wavelet Transformation (HWT). In the case of dimension (x) they have x = {[32 × 32], [48 × 48], [128 × 128]}, and for simplicity, x = {32, 48, 128} is employed. All of these measurements refer to the scaling quality of the input photos. CL and PL have y = 2, 3: each of the two is obtained from CNN, and y = 2 indicates that CNN has two pairs of folded levels and pooling sections. A.Kumar et. al. (2018) proposed the method that extract CNN and the LSTM architecture are used to identify locally and globally characteristics. This framework mainly focusses on to solve the problem of script identification of video scripts and text images. Initially they convert image into patches and feed them into the proposed method. After that they did to get feature points, multiply the matching CNN patch by patch. Global characteristics are also extracted from the LSTM's final last convolution layer. Fusion technique was employed to weight local and global features of individual patches. Four public SIW-13, CVSI2015, ICDAR-17, and MLe2e databases were used to test the approaches. A quick summary of the suggested scheme is shown in Fig.5.

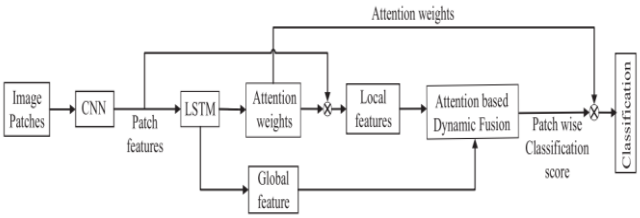

Figure 5: CNN LSTM Framework Bhunia A.K et al. (2018).

The end-to-end framework depicted in Figure 5 is made up of 3 steps. It starts by extracting precise translationally invariant image attributes using a stacked folding layer structure. Feature vectors with various dimensions are generated using CNN levels. To take use of the spatial dependencies present in the text script pictures, these vectors are incorporated into the LSTM layer.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

The attentiveness network comes next, following by a Softmax layer to keep the patch weights in check. The attention model was chosen in order to give significance to the most significant elements. The local characteristics for the individual patches are computed by adding these concentration weights by the retrieved CNN exclusive use with the patches. A perfectly alright depiction of the printed text can be found in these local characteristics. The worldwide feature is also taken from the last current block of the LSTM unit in order to acquire the techniques of analysis of these images. Furthermore, we combined the locally and globally features gained in the second step using dynamic media exposure weighing. Finally, the categorization ratings for each patches are compared to a completely linked level. To acquire the final probabilistic model across the categories, the critical in assessing incorporates the sum as a function of the attention of these classification evaluations per patches. It gets over the restriction of summing by components, which treats all adjustments identically.

DATA ANALYSIS AND RESULTS

In above section we have been summarized the latest stat-of-art techniques for multilingual script identification on natural images. The results shows quite satisfy on particular datasets. It's important to know the behavior, advantages and limitations of state-of-art techniques. In Table 1, shows comparison of latest work on script identification on multilingual text. As CNN-patch is a base method for B. Shi (2015), so they also use CNN-Patch for SIW-10, got low percentage results, as CNN is incapable of dealing with input of any scale. To address this issue, B. Shil et al. (2015) resize images in the start with aspect- ratio, then using sliding windows make same sizes of patches after sampling for training phase. They trained with learning period rate set to 0.01 and the velocity set to 0.9, the network was trained using stochastic gradient descent (SGD) P. Krishnan et al. (2014). Locality Constraints linear coding Y. LeCun (1998), is a bag-of-words framework, widely used for image classification. LLC have little better results as compared to CNN patch on SIW-10 dataset but not reach to proposed algorithm that is MSPN.

Table1: Ccomparison Results

Researchers	Features	Classifiers	Database	Results
B. Shi et al. (2015)	Hierarchical features by CNN	MSPN	SIW-10	94.4%
A.Ul-Hassan et al. (2015)	Sequence of Class-labels	LSTM	English-Greek text	98.90%
A. P. Singh et al. (2015)	Density of the pixels	RNN+CTC	12-Indic Script	95.5%
A. Nicolaou et al. (2016)	(SRS)-LBP variant of LBP histogram features	Deep Multi-Layer Perceptron (MLP)	CVSI 2015+SIW-10+SIW-13	97.3%, 94.6%
A.K. Singh et al. (2016)	RNN	SVM	CVSI 2015+ ILST	96.70%+ 88.67%
L. Gomez et al. (2016)	CNN	NNBN	CVSI 2015 +MLe2e	97.91% + 91.12%
Lu, Liqiong, et al. (2019)	Local and Global CNN	ResNet-20	CVSI 2015+MLe2e+SIW-13+ICDAR	98.3% + 95.8% +96.1% +93.2%

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

For comparison with another algorithm GLCM for short (Gray-Level Co-occurrence Matrix Features) J. Hochberg (1997) is used but the result is not prominent as MSPN have. AQs some Because many languages overlap a significant number of alphabet letters, factor plays an important between multiple pairings of scripts, such as Chinese and Japanese, are common. Without meaning, many languages are much more complicated. that's why they have not use this type of data into MSPN framework. Comparison shows the performance of GLCM A. Busch (2015) is very low due to a complete lack of ability to differentiate between scripts with similar looks Due to the usage of Pouch models in feature extraction, LLC beats CNN-Patch, but MSPN surpasses LLC by a wide margin as deep models are stronger learner. This paper A. Ul-Hasan (2015) explains a novel process for recognising numerous scripts in multinational texts significant sequence sequencing. It works on bilingual data and show the encouraging the identification results. In A. K. Singh et al. (2015) developed an efficient method for script identification on the overall average of 12 different languages in the Indian multilingual dataset is 95.55 percent, which is higher than the 94.44 percent given by P. B. Pati et al. (2008) and A. Nicolaou (2016) suggest the identification method based on traditional way of feature extraction that is hand-crafted texture features used and an artificial neural network for classification. Proposed method got the prominent results but still need chasing for results of stat of the art techniques. A.k. Singh et al (2016) Consider the problems of script identification on wild images. This approach obtained bag of mid-level features from script images. Novelty of L. Gomez et al. (2016) is the representation of for the classifier, convolutional layers and Naive-Bayes Nearest Neighbour are used. It exhibits the potential of script recognition in wild scene photos, which aids in complete multilingual end-to-end scenario text comprehension.

CONCLUSION

A bird's eye perspective of multilingual character recognition is presented in this work stat-of-art techniques. Multilingual script identification has an important role in OCR research area in the natural and wild images text recognition system. Researchers have tried to detect scripts by identifying structural characteristics or deriving aesthetic characteristics of input text images in the system. Briefly explains methods for comparison analysis, characteristics and learners were employed for character recognition, the issues involved in character recognition, quantitative analysis, and characteristics and learners. In reality, research is matured in the area of script identification in documents; however, the same accuracy level is not met with multilingual script identification on natural scenes. The problem for multilingual script identification remains an open problem very much due to noisy images, stylish and variety of fonts in Scene text images, images have different aspect ratios, complex background etc. To handle these types of problems still there is a necessity of new techniques for more salient feature extraction and classifiers. Therefore, it is required to propose and develop new strategies to improve accuracy and overall accuracies for multilingual script identification on natural scenes.

ACKNOWLEDGMENT

Should be brief, mention people directly involved with the research project (i.e. advisors, sponsors, funding agencies, colleagues, technicians or statistical helper who have supported your work).

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

REFERENCES

[1] Ghosh, D., Dube, T., & Shivaprasad, A. (2010). Script recognition—a review. IEEE Transactions on pattern analysis and machine intelligence, 32(12), 2142-2161.

[2] Singh, A. K., & Jawahar, C. V. (2015, August). Can RNNs reliably separate script and language at word and line level? In 2015 13th International Conference on Document Analysis and Recognition (ICDAR) (pp. 976-980). IEEE.

[3] Chanda, S., Pal, S., Franke, K., & Pal, U. (2009, July). Two-stage approach for word-wise script identification. In *2009 10th International Conference on Document Analysis and Recognition* (pp. 926-930). IEEE.

[4] Basu, S., Das, N., Sarkar, R., Kundu, M., Nasipuri, M., & Basu, D. K. (2010). A novel framework for automatic sorting of postal documents with multi-script address blocks. *Pattern Recognition*, *43*(10), 3507-3521.

[5] Bazzi, I., Schwartz, R., & Makhoul, J. (1999). An omnifont open-vocabulary OCR system for English and Arabic. *IEEE Transactions on pattern analysis and machine intelligence*, *21*(6), 495-504.

[6] Gomez, L., & Karatzas, D. (2013, August). Multi-script text extraction from natural scenes. In *2013 12th International Conference on Document Analysis and Recognition* (pp. 467-471). IEEE.

[7] Li, L. J., Socher, R., & Fei-Fei, L. (2009, June). Towards total scene understanding: Classification, annotation and segmentation in an automatic framework. In *2009 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 2036-2043). IEEE.

[8] Cour, T., Jordan, C., Miltsakaki, E., & Taskar, B. (2008, October). Movie/script: Alignment and parsing of video and text transcription. In *European Conference on Computer Vision* (pp. 158-171). Springer, Berlin, Heidelberg.

[9] Alabau, V., Sanchis, A., & Casacuberta, F. (2014). Improving on-line handwritten recognition in interactive machine translation. *Pattern Recognition*, *47*(3), 1217-1228.

[10] Toselli, A. H., Romero, V., Pastor, M., & Vidal, E. (2010). Multimodal interactive transcription of text images. *Pattern Recognition*, *43*(5), 1814-1825.

[11] Tan, T. N. (1998). Rotation invariant texture features and their use in automatic script identification. *IEEE Transactions on pattern analysis and machine intelligence*, *20*(7), 751-756.

[12] Hochberg, J., Kelly, P., Thomas, T., & Kerns, L. (1997). Automatic script identification from document images using cluster-based templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *19*(2), 176-181.

[13] Busch, A., Boles, W. W., & Sridharan, S. (2005). Texture for script identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(11), 1720-1732.

[14] Joshi, G. D., Garg, S., & Sivaswamy, J. (2007). A generalised framework for script identification. *International Journal of Document Analysis and Recognition (IJDAR)*, *10*(2), 55-68.

[15] Phan, T. Q., Shivakumara, P., Ding, Z., Lu, S., & Tan, C. L. (2011, September). Video script identification based on text lines. In *2011 International Conference on Document Analysis and Recognition* (pp. 1240-1244). IEEE.

[16] Zhao, D., Shivakumara, P., Lu, S., & Tan, C. L. (2012, March). New spatial-gradient-features for video script identification. In *2012 10th IAPR international workshop on document analysis systems* (pp. 38-42). IEEE.

[17] Shi, B., Bai, X., & Yao, C. (2016). Script identification in the wild via discriminative convolutional neural network. *Pattern Recognition*, *52*, 448-458.

[18] Shi, B., Bai, X., & Yao, C. (2016). Script identification in the wild via discriminative convolutional neural network. *Pattern Recognition*, *52*, 448-458.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

[19] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
[20] Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., & Gong, Y. (2010, June). Locality-constrained linear coding for image classification. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 3360-3367). IEEE.
[21] Ul-Hasan, A., Afzal, M. Z., Shafait, F., Liwicki, M., & Breuel, T. M. (2015, August). A sequence learning approach for multiple script identification. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 1046-1050). IEEE.
[22] "OCRopus – Open Source Document Analysis and OCR System." [Online]. Available: https://github.com/tmbdev/ocropy.
[23] Breuel, T. M., Ul-Hasan, A., Al-Azawi, M. A., & Shafait, F. (2013, August). High-performance OCR for printed English and Fraktur using LSTM networks. In *2013 12th international conference on document analysis and recognition* (pp. 683-687). IEEE.
[24] Singh, A. K., & Jawahar, C. V. (2015, August). Can RNNs reliably separate script and language at word and line level?. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 976-980). IEEE.
[25] Krishnan, P., Sankaran, N., Singh, A. K., & Jawahar, C. V. (2014, April). Towards a robust OCR system for Indic scripts. In *2014 11th IAPR International Workshop on Document Analysis Systems* (pp. 141-145). IEEE.
[26] Rath, T. M., & Manmatha, R. (2003, August). Features for word spotting in historical manuscripts. In *Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings.* (pp. 218-222). IEEE.
[27] Graves, A., Fernández, S., Gomez, F., & Schmidhuber, J. (2006, June). Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks. In *Proceedings of the 23rd international conference on Machine learning* (pp. 369-376).
[28] Nicolaou, A., Bagdanov, A. D., Gomez, L., & Karatzas, D. (2016, April). Visual script and language identification. In *2016 12th IAPR workshop on document analysis systems (DAS)* (pp. 393-398). IEEE.
[29] Nicolaou, A., Bagdanov, A. D., Liwicki, M., & Karatzas, D. (2015, August). Sparse radial sampling LBP for writer identification. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (pp. 716-720). IEEE.
[30] Sharma, N., Mandal, R., Sharma, R., Pal, U., & Blumenstein, M. (2015, August). ICDAR2015 competition on video script identification (CVSI 2015). In *2015 13th international conference on document analysis and recognition (ICDAR)* (pp. 1196-1200). IEEE.
[31] Singh, A. K., Mishra, A., Dabral, P., & Jawahar, C. V. (2016, April). A simple and effective solution for script identification in the wild. In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)* (pp. 428-433). IEEE.
[32] Fernando, B., Fromont, E., & Tuytelaars, T. (2014). Mining mid-level features for image classification. *International Journal of Computer Vision*, *108*(3), 186-203.
[33] Juneja, M., Vedaldi, A., Jawahar, C. V., & Zisserman, A. (2013). Blocks that shout: Distinctive parts for scene classification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 923-930).
[34] Boureau, Y. L., Bach, F., LeCun, Y., & Ponce, J. (2010, June). Learning mid-level features for recognition. In *2010 IEEE computer society conference on computer vision and pattern recognition* (pp. 2559-2566). IEEE.
[35] Pati, P. B., & Ramakrishnan, A. G. (2008). Word level multi-script identification. *Pattern Recognition Letters*, *29*(9), 1218-1229.
[36] Ojala, T., Pietikainen, M., & Maenpaa, T. (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(7), 971-987.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

[37] Gomez, L., & Karatzas, D. (2016, April). A fine-grained approach to scene text script identification. In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)* (pp. 192-197). IEEE.

[38] Coates, A., Ng, A., & Lee, H. (2011, June). An analysis of single-layer networks in unsupervised feature learning. In Proceedings of the fourteenth international conference on artificial intelligence and statistics (pp. 215-223). JMLR Workshop and Conference Proceedings.

[39] Boiman, O., Shechtman, E., & Irani, M. (2008, June). In defense of nearest-neighbor based image classification. In 2008 IEEE conference on computer vision and pattern recognition (pp. 1-8). IEEE.

[40] Yao, B., Bradski, G., & Fei-Fei, L. (2012, June). A codebook-free and annotation-free approach for fine-grained image categorization. In *2012 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3466-3473). IEEE.

[41] Krause, J., Gebru, T., Deng, J., Li, L. J., & Fei-Fei, L. (2014, August). Learning features and parts for fine-grained recognition. In *2014 22nd International Conference on Pattern Recognition* (pp. 26-33). IEEE.

[42] Lu, L., Yi, Y., Huang, F., Wang, K., & Wang, Q. (2019). Integrating local CNN and global CNN for script identification in natural scene images. *IEEE Access*, *7*, 52669-52679.

[43] Ukil, S., Ghosh, S., Obaidullah, S. M., Santosh, K. C., Roy, K., & Das, N. (2020, January). Deep learning for word-level handwritten Indic script identification. In *International conference on recent trends in image processing and pattern recognition* (pp. 499-510). Springer, Singapore.

[44] Kerautret, B., Colom, M., Lopresti, D., Monasse, P., & Talbot, H. (Eds.). (2019). *Reproducible Research in Pattern Recognition: Second International Workshop, RRPR 2018, Beijing, China, August 20, 2018, Revised Selected Papers* (Vol. 11455). Springer.

[45] Roy, P. P., Bhunia, A. K., & Pal, U. (2018). Date-field retrieval in scene image and video frames using text enhancement and shape coding. *Neurocomputing*, *274*, 37-49.

[46] Obaidullah, S. M., Santosh, K. C., Das, N., Halder, C., & Roy, K. (2018). Handwritten Indic script identification in multi-script document images: a survey. *International Journal of Pattern Recognition and Artificial Intelligence*, *32*(10), 1856012.

BIOGRAPHY

Kiran Perveen received the B.S. degree in Computer Science from University of Agriculture, Faisalabad, Pakistan in 2017. She received MS degree in Computer Science from COMSATS Institute of Information Technology, Islamabad, Pakistan in 2021. She is currently working as Research Assistant & Embedded System Engineer at CACSYS Engineering. Her current research interests include Artificial Intelligence, Machine learning, Image Processing and Embedded Systems.

Logical Creations Education Research Institute
LC INTERNATIONAL JOURNAL OF STEM
E-ISSN: 2708-7123
Web: www.lcjstem.com | Email: editor@lcjstem.com
Volume-03 | Issue-01 | March-2022

Rukhsana Perveen received the B.S. degree in Electrical Engineering from University of Gujrat (UOG), Pakistan in 2013. She received MS degree in Electrical Engineering from National University of Computer and Emerging Sciences (FAST-NU), Pakistan in 2019. She is currently working as PCB & Embedded System Designer at CACSYS Engineering. Her current research interests include Power System, Robotics and Embedded Systems.

Dr. Awais Yasin received the B.S. degree from the University of Engineering and Technology, Taxila, and the M.S. and Ph.D. degrees in the field of robotics from the Beijing Institute of Technology, Beijing, China, in 2013. He has a rich industrial experience of more than 16 years serving different research organizations of Pakistan. He is currently the Director of the Virtual Reality and Machine Learning Laboratory, UET Taxila, an Associate Professor with NUTECH, Islamabad, and a member of Industrial Advisory Board (IAB) for more than ten different universities of Pakistan. His research and teaching interests include robotics, machine learning, virtual reality, drive simulation, and intelligent transportation systems. He was honored to win the World Cup Award in the field VR Cloud Programming held in Tokyo, Japan, in 2018.