SCRIPT IDENTIFICATION: A REVIEW

[1]Ranjitha C,R, [2]Reema Mathew A, [3]Lekshmy S
[1,2,3]Dept.of electronics and communication, Vimaljyothi engineering college, Kannur,kerala,India
[1]ranjithacr12@gmail.com, [2]reemamathew@vjec.ac.in, [3]lekshmyhari@vjec.ac.in

Abstract: In a multilingual, robust learning environment, identifying a script in the field is very important. Textual identification is an important task, especially in India, where there are 13 different texts for 22 languages. Text filtering, automatic translation, OCR (Optical Character Recognition) and text location identification are the main applications for script identification. In recent years, with the widespread use of the Internet and automated text processing around the world, scripting techniques have become increasingly important in the field of pattern recognition. Script Identification refers to techniques for distinguishing different texts into multilingual and graphic texts.

I.INTRODUCTION

Every human core of the population has a collection of languages which belong to that country and are considered to be its inherent characteristic. The root of human languages has been the topic of intellectual discussion for many decades. Even after so much study, there was no agreement on definite origin. Similarly, no agreement was reached at the age of the human language. This problem is made more complex by the fact that there is a deficit in direct undeviating facts. As a result, researchers seeking to discover and investigate the origin and genesis of languages must draw inferences from other forms of data and information such as archaeological evidence, language learning hypotheses, fossil records, current linguistic diversity, and by similarities and analogies between the human communication system and the communication systems used by animals. For communication of messages in a language, a writing method is popularly defined as a systematic, structured, and routine process for storing and transmitting text. This is achieved by using a series of symbols widely referred to as characters, for visual encoding (writing) and decoding (reading). Collectively, the list of these characters is referred to a script. Collections of these characters typically contain numbers and letters. The attributes of writing systems can be broadly categorized into:

1. Alphabets: includes a standard set of letters consisting of vowels and consonants which encode, on the basis of the general law, that the letters reflect simple, significant sounds which are the phonemes of the spoken language.
2. Syllabaries: Generally, the syllabaries here correlate a syllable to a sign (these are usually a pair or group of phonemes and these are used as units for building words).
3. Logography: Here the character represents each unit of sentences, phrase or morpheme. They can be in groups or two groups of characters.

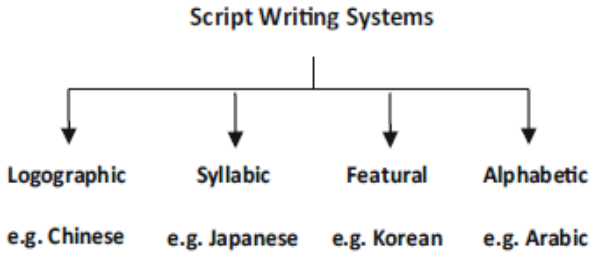

Fig. 1 Common script writing systems

Script Identification is intended to predict the script of a given text, which has a very important role in multilingual programming. Under several areas, it is necessary to determine which language model should be used for further identification or textual recognition. Pre-paper work, handwriting and video overlay, in which texts have a clear layout and clear context, have achieved great efficiency. But in the case of identifying Scene Text Script, which extends the application to many fields such as image Comprehension, other problems arise, such as complex content, different text types and different sounds, and so on. Our work focuses

on the text of the forum, taking into account the following challenges:

• Some scripts have minor variations, e.g. Russian and English, Tamil and Telugu, which share a wide group of characters. Divorce is actually a problem of good segregation that includes racist elements.

• The overlapping text images have contradictory dimensions, which makes it important to find the right way to model them during the bulk-based training process.

The first problem is important in identifying the script, where the bottleneck arises primarily from the same family scripts, which share certain identical characters. Local racist factors are therefore often overlooked. Almost all functions focus on collecting sensitive features without compressing obsolete features that act as audio.Other functions follow the integration of deeper aspects of problem solving. There was a training section with many stages and counting as a result of the merger.

II. GENERAL BACKGROUND

Mental speaking has been a very difficult task for the people, and they have tried to communicate internally for thousands of years, and it is considered that they have achieved this skill. The use of writing is one of the traditional and fundamental forms of this representation. Writing is considered the art of expressing speech in a figurative, expressive and visual way. Writing has become an important means of communication and communication that expresses emotions and often language by capturing or writing with the help of symbols and symbols.Writing methods have evolved from counting and tracking time over long periods known as calendars. It is also possible that it originated in ancient Egypt and Mesoamerica in need of political and historical political writings. Repetition of writing programs took place automatically by the emergence of computer programs. As they used to write on paper, people are able to write in machine-based applications. In these computer programs, records containing specific documents are edited, stored, and used in a variety of languages. Computer-based applications receive hundreds of written languages using standard text. Different texts are used to write different languages. Manuscript or paper is often referred to as the "Script". Sometimes a script is also called a handwritten text / page / document. The set of letters and characters used for writing is usually related. The most important and important aspect of each text is simply this text.

III. SCRIPT IDENTIFICATION SYSTEMS

Automated text management and analysis has been developing as an important field of research and development over the past few decades. The acquisition of script type is an important and fundamental aspect of the management and analysis of document documents. The text contained in a written document is regarded as text. Texts are basically written using codes. Script recognition is the way in which a single text image sees a corresponding script for language. The various script identification programs are:

- Scene Text Script Identification with Convolutional Recurrent Neural Networks
- Sequence-to-label script identification for multilingual optical character recognition.
- attention based Convolutional-LSTM network
- convolutional triplets for script identification in scene text
- Discriminative convolutional neural network
- script identification method based on hand-crafted texture features and an artificial neural network
- fully differentiable method for multilanguage scene text localization and recognition
- Integrating local CNN and global CNN

A. Scene Text Script Identification with Convolutional Recurrent Neural Networks

This approach uses its combination of the convolutional neural network (CNN) and the recurrent neural network (RNN). CNN produces rich image presentations, and RNN effectively analyzes long-term local dependencies. CNN is one type of neural network that is enhanced by the inclusion of a convolution function. CNNs have a much more complex structure than conventional presentations, involving several layers of non-linear elements. With an integrated structure, CNN's capabilities can be

controlled with a variety of variations and widths. Features produced by CNN are read directly from manual data, with minimal size and discriminatory details. Duplicate neural networks Feedforward neural networks have edges that connect the immediate steps, presenting the view of time to the network.

The end-to-end network construction is primarily about three things. In order to obtain accurate image translation presentations, we use the convolutional layers layer as the first part of our model. In the aftermath of convolutional layers, the RNN layers take the conflicting feature length maps generated by the CNN layers as a means of applying local dependence on script images. To integrate the release of RNN layers, we include the integration layer behind the RNN layers. The third component is a fully integrated layer that is widely used for segmentation problems. As with the installation of the entire model, the text images are saved so that they have a consistent height while having different widths to maintain the same dimensions. Consisting of three elements, the exit model can still be trained end-to-end spread due to the unique properties of all of these items.

B.Sequence to label script identification for multilingual ocr

Figure 2 shows our approach to using the multilingual OCR. In the first step, the text lines are removed from the image. Details of the structure of the sections or sections can be provided in this section. It can also provide style details, as described in., As a bold or italic style. For the first 30 scripts, script recognition selects an OCR model that is specific to a particular script. OCR script-specific model and create unicode image line text. Finally, recognition results continue to improve structural analysis.

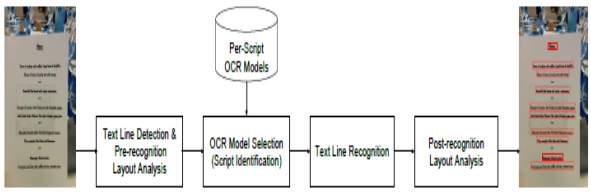

Fig. 2. Multilingual OCR system with line-level script identification.

For instance, post OCR errors in the initial layout are resolved by using low confidence recognition as a signal that an initial text box was not actually text.

C.Attention based convolutional LSTM network

This approach incorporates CNN's in-depth design into image patches to extract their presentations and ultimately give them LSTM Afterwards to provide a number of those most important features, using a method to pay attention to the weight calculation of patches. The clever duplication of vectors by the CNN elements extracted from these sensory devices reveals the local features of the individual patches, while the global element is found in the last cell of the LSTM. Local features include well-crafted details while the global feature captures full representation of text images. Using powerful scales, domestic and global features are finally integrated. This approach incorporates dynamic measurements based on attention to determine which, in terms of their value, can provide the greatest measurement between a land element and a spatial element. To achieve the scope of each patch division, a fully integrated layer is used at the top.

Depending on their relationship and integrated representation, a strong measure of domestic and global symbols is used. Together, two different types of features will effectively reduce the limit of each feature. Final divisions require a summary of school divisions wisely. It overcomes the smarter summary that gives all episodes equal value.

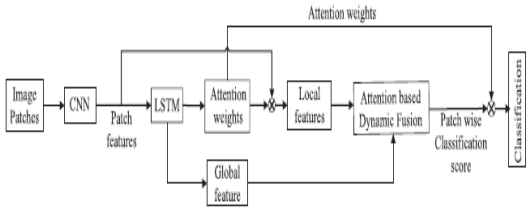

Fig.3.Attention based convolutional LSTM network

D.convolutional triplets for script identification in scene text

The main focus of this work is to improve the recognition of the occurrence of many common characters of characters and patterns of characters. This method uses three-dimensional place markers of bag-of-view-words (BoVW) code word encoding, inspired by a visual bag bag where images are represented by successive words, as well as a discriminatory rendering of triplets image patches. Using a combination of local definitions, we aim to maximize the power of representation of relevant codewords in the

BoVW model. Weak definitions benefit from a combination of other definitions, which leads to the definition of triplet descriptor.

This program uses script recognition at the line of text line, meaning that a single-label label in the classification process is applied to each line of text. Pre-segmented text images are converted to a gray space and their difference is usually in the range of [0, 1]. By extracting square dots from text-line images in a window-sized training window in two-dimensional, high-line text height, and two-thirds of the text-line length, we created a group of image dot training. The extracted dots were enlarged to a fixed size of 32 to 32 pixels and used to train CNN to separate fragments of individual text.

E. Discrimination of the convolutional neural network

The basic idea is to incorporate in-depth features and intermediate presentations into a deeper model that can be trained globally. Specifically, the in-depth feature map collection is first collected from input images with a pre-trained CNN model, in which the in-depth features of the area are heavily collected. After that in order to read a collection of racist patterns based on those local factors, a racist collection is made. Medium presentation is achieved by coding local features based on discriminated patterns learned (codbook). Finally, in a deep network, intermediate level representation and in-depth features are equally configured. With the help of this fine-tuning process, a well-prepared model, called the Discriminative Convolutional Neural Network (DisCNN), is able to accurately detect subtle differences between texts that are difficult to distinguish.

Representation between discriminatory levels based on in-depth features is provided by script recognition functions, unlike other methods that focus on sewing, editing or analyzing a related component. Here intermediate level representation and deep work removal can be combined with a deeper model and developed collaboratively. The suggested method is not limited to wild script recognition, and also applies to video and paper script identification.

F. script identification method based on hand-crafted texture features and an artificial neural network

This paper deals with the production of script or language Identification in a variety of ways, such as video text, status text or handwritten text, and the use of a text-based approach and a fully integrated neural network. We show that the separation of NN over elements collected from the first layer of the deep neural network is similar to or exceeds the deep network partition. and the functional use of the neural network as a studied metric to produce flexible differentiation.

This approach contains a pre-processing phase followed by the removal of LBP (binary local pattern) features and Artificial Neural Network (ANN) training in these features. The middle layers of the ANN are then used as a production model for segmentation purposes. Before transferring images to LBP transformation, each image is edited independently. Since the LBP conversion is applied to a single channel image, instead of illumination, a major component of all pixel colors was selected to enhance the apparent contrast that can be caused by light. To have a consistent LBP encoding between images with a bright black background and images with a black background in a bright background, whenever the central band becomes darker than the image size, the image is investigated. It is assumed that more front pixels will be present in the band between between 25% and 75% of the image width.

G. Fully differentiable method for multilanguage scene text localization and recognition

E2E-MLT, an end-to-end FCN system with decentralized sharing layers for multilingual text. E2EMLT enables multilingual text, text recognition and script identification using a single fully integrated network. This method is suitable for the following languages: Arabic, Bangla, Chinese, Japanese, Korean, Latin and is able to understand 7500 characters (compared to less than 100 English characters and does not use any word dictionary). E2E-MLT receives local text creation from a multilingual text image environment, and generates text writing and script classes in each of the acquired regions.

Fully convolutional networks (FCN), is a network that does not have "Dense" layers (as in traditional CNNs) but instead contains 1x1

combinations that perform the function of fully connected layers (Dense layers).

H. Inteegrating local CNN and global CNN

Combining local CNN and global CNN both are based on ResNet-20 script recognition. We are starting to get more patches and split images depending on the size of the images. Afterwards, these tags and split images are used for CNN Home and Global CNN training inputs, respectively. Finally, to obtain final results, the Adaboost algorithm is used to combine Local CNN results with Global CNN for decisionlevel fusion. To take advantage of such a strategy, CNN Local makes full use of local image features, effectively revealing subtle differences between hard-to-distinguish scripts such as English, Greek, and Russian. In addition, Global CNN mines global image features to improve the accuracy of script identification.

IV.DATASETS

A few key reasons allow for the need for additional data stocks: i) Many deep learning methods are driven by data. High-level data stocks are important and crucial to training a good text identifier. ii) New data sets often represent indicators of future work, such as anonymous text recognition, unconventional text recognition, unchecked or underestimated text recognition, and large-scale text recognition.

Depending on the type of data set, we classify standard layout data sets into two categories: automated data sets and actual datasets. In particular, virtual datasets include standard Latin databases, non-standard Latin databases and multilingual datasets.

A.Synthetic datasets

The most deep learning algorithms rely on sufficient data. However, the actual data sets are too small for training the most accurate group text identifier, because contain only thousands of data samples. In addition, manually collecting and interpreting large amounts of real data will involve significant efforts and resources.

- Synth90k:The Synth90k database contains 9 million sample images of text from some of the English 90k words. Names are given to natural images with random modifications and effects, such as random fonts, colors, blurring and sounds. The Synth90k database can mimic the distribution of graphic text images and can be used instead real world data training deep learning algorithms for data. Besides, the whole picture is described by the true name of the earth.
- SynthText:The SynthText database contains 800000 images with six million synthetic text forms. Like the Synth90k database generation, a text sample is provided using a randomly selected font and converted according to local location. Moreover, each image is defined by the true name of the world.
- Verisimilar Synthesis:The Verisimilar Synthesis database contains 5 million images for text modeling. Given background images and source text, the semantic map and saliency map are first determined by compiling and identifying logical and appropriate areas for text embedding. The color, brightness, and layout of the text are further determined by the color, brightness, and textures surrounding the embedded areas within the background image.
- Unreal Text:The UnrealText database contains 600K artificial images with 12 million text formats. Text conditions are considered as planar polygon meshes with pre-text areas loaded as textures. These meshes are set in exact positions in the 3D world, and provided with the whole scene. The same font set from Google Fonts3 and the same corpus, i.e., Newsgroup20, are used as SynthText does.

Fig. 4: Synthetic sample images of text from Synth90k and SynthText datasets.

B.Realistic Datasets

Most current databases contain only thousands of text for example text. Thus, for STR, logical data sets are commonly used to test recognition algorithms under real-world conditions. Next, we will document and briefly

describe existing databases: standard Latin databases, non-standard Latin data sets, and multilingual databases.

- Regular Latin Datasets:In standard Latin databases, most text types are front and horizontal, with a small portion of them distorted.
- ICDAR 2011 (IC11):The IC11 database contains 485 images. This is an extension of the database used for ICDAR 2003 text acquisition competitions.
- ICDAR 2013 (IC13):The IC13 database contains 561: 420 training images and 141 test images. It inherits data from IC03 databases and expands it with new images.
- Unusual Latin Datasets:In datasets of random bench shares, most text cases have low resolution, distorted view, or curved.
- COCO-Text:The COCO-Text database contains 63686 images.

V.APPLICATIONS

The text, as the most important part of communication and exploration of the world, enriches our lives. Many applications for text recognition in various industries and in our daily lives include text comprehension, information extraction, visual text query answers, image classification, archival archiving, automated geocoding systems, multimedia retrieval, water meter reading, regular text to speech devices. Table 1 introduces some independent applications.

Applications of scene script identification		
Education industry	Medical health industry	others
Automatic grading of examination papers	Digitization of medical records	Smart robots
Augmented reality industry	Big data industry	Insurance industry
Real time street view AR translation	Massive extraction of text from images	Information extraction
Intelligent transportation system	Smart government	Image classification

Table1.Examples of various applications of scene text identification

CONCLUSION

In this review paper, a review of the various script identification strategies for global and national contexts was presented. Lastly, a collection of datasets and applications is included.

REFERENCES

[1] L. Gomez, A. Nicolaou, and D. Karatzas, “Improving patch-based scene text script identification with ensembles of conjoined networks,” Pattern Recognition, vol. 67, pp. 85–96, 2017.

[2] J. Mei, L. Dai, B. Shi, and X. Bai, “Scene text script identification with convolutional recurrent neural networks,” in 2016 23rd InternationalConferenceonPatternRecognition(ICPR). IEEE,2016, pp. 4053–4058.

[3] B. Shi, X. Bai, and C. Yao, “An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition,” IEEE transactions on pattern analysis and machine intelligence, vol. 39, no. 11, pp. 2298–2304, 2017.

[4] Y. Fujii, K. Driesen, J. Baccash, A. Hurst, and A. C. Popat, “Sequence-to-label script identification for multilingual ocr,” in Document Analysis and Recognition (ICDAR), 2017 14th IAPR International Conference on, vol. 1. IEEE, 2017, pp. 161–168.

[5] A. K. Bhunia, A. Konwer, A. K. Bhunia, A. Bhowmick, P. P. Roy, and U. Pal, “Script identification in natural scene image and videoframesusinganattentionbasedconvolutional-lstmnetwork,” Pattern Recognition, vol. 85, pp. 172–184, 2019.

[6] Y.Wang,V.I.Morariu,andL.S.Davis,“Learningadiscriminative filter bank within a cnn for fine-grained recognition,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp. 4148–4157.

[7] H. Zheng, J. Fu, T. Mei, and J. Luo, “Learning multi-attention convolutional neural network for fine-grained image recognition,” in Int. Conf. on Computer Vision, vol. 6, 2017.

[8] B. Shi, X. Bai, and C. Yao, “Script identification in the wild via discriminativeconvolutionalneuralnetwork,”PatternRecognition, vol. 52, pp. 448–458, 2016.

[9] N. Nayef, F. Yin, I. Bizid, H. Choi, Y. Feng, D. Karatzas, Z. Luo, U. Pal, C. Rigaud, J. Chazalon et al., "Icdar2017 robust reading challenge on multi-lingual scene text detection and script identificationrrcmlt,"inDocumentAnalysisandRecognition(ICDAR),201714thIAPRInternationalConferenceon,vol.1. IEEE, 2017, pp. 1454–1459.

[10] X. Zhang, H. Xiong, W. Zhou, W. Lin, and Q. Tian, "Picking deep filterresponsesforfine-grainedimagerecognition,"inProceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 1134–1142.

[11] Y. Patel, M. Bu˘sta, and J. Matas, "E2e-mlt-an unconstrained end-to-end method for multi-language scene text," arXiv preprint arXiv:1801.09919, 2018.

[12] L. Gomez and D. Karatzas, "A fine-grained approach to scene text script identification," in Document Analysis Systems (DAS), 2016 12th IAPR Workshop on. IEEE, 2016, pp. 192–197.

[13] M.Cordts,M.Omran,S.Ramos,T.Rehfeld, M.Enzweiler,R.Benenson,U.Franke,S.Roth,and B.Schiele,"Thecityscapesdataset for semantic urban scene understanding," in Proceedings of the IEEEconferenceoncomputervisionandpatternrecognition,2016, pp. 3213–3223.

[14] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, ''Reading text in the wild with convolutional neural networks,'' Int. J. Comput. Vis., vol. 116, no. 1, pp. 1–20, 2016.

[15] W. Lu, H. Sun, J. Chu, X. Huang, and J. Yu, ''A novel approach for video text detection and recognition based on a corner response feature map and transferred deep convolutional neural network,'' IEEE Access, vol. 6, pp. 40198–40211, 2018.

[16] W. Lu, H. Sun, J. Chu, X. Huang, and J. Yu, ''A novel approach for video text detection and recognition based on a corner response feature map and transferred deep convolutional neural network,'' IEEE Access, vol. 6, pp. 40198–40211, 2018.

[17] Q. Ye and D. Doermann, ''Text detection and recognition in imagery: A survey,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 37, no. 7, pp. 1480–1500, Jul. 2015.

[18] D. Ghosh, T. Dube, and A. Shivaprasad, ''Script recognition—A review,'' IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 12, pp. 2142–2161, Dec. 2010.

[19] T. Q. Phan, P. Shivakumara, Z. Ding, S. Lu, and C. L. Tan, ''Video script identification based on text lines,'' presented at the 11th Int. Conf. Document Anal. Recognit., Beijing, China, Sep. 2011.

[20] K. Simonyan and A. Zisserman, ''Very deep convolutional networks for large-scale image recognition,'' presented at the Int. Conf. Learn. Represent., San Diego, CA, USA, May 2015.

[21]] N. Sharm, R. Mandal, R. Sharma, U. Pal, and M. Blumenstein, ''ICDAR2015 competition on video script identification (CVSI 2015),'' presentedatthe13thInt.Conf.DocumentAnal.Recognit.,Nancy,France, Aug. 2015.

[22] N. Nayef et al., ''ICDAR2017 robust reading challenge on multilingual scene text detection and script identification-RRC-MLT,'' presented at the 14th Int. Conf. Document Anal. Recognit., Kyoto, Japan, Nov. 2017.

[23] K. Ubul, G. Tursun, A. Aysa, D. Impedovo, and G. Pirlo, ''Script identification of multi-script documents: A survey,'' IEEE Access, vol. 5, pp. 6546–6559, Mar. 2017.

[24] N. Sharma, S. Chanda, U. Pal, and M. Blumenstein, ''Word-wise script identification from video frames,'' presented at the 12th Int. Conf. Document Anal. Recognit., Washington, DC, USA, Aug. 201

[25] N. Sharma, U. Pal, and M. Blumenstein, ''A study on word-level multiscript identification from video frames,'' presented at the Int. Joint Conf. Neural Netw., Beijing, China, Jul. 2014.

[26] L. Gómez and D. Karatzas, ''A fine-grained approach to scene text script identification,'' presented at the 12th IAPR Workshop Document Anal. Syst., Santorini, Greece, Apr. 2016.

[27] M. Verma, N. Sood, P. P. Roy, and B. Raman, ''Script identification in natural scene images: A dataset and texture-feature based performance evaluation,'' presented at the Int. Conf. Comput. Vis. Image Process., vol. 460. Singapore, Springer, 2017.

[28] A. Singh, A. Mishra, P. Dabral, and C. Jawahar, ‘‘A simple and effective solution for script identification in the wild,’’ presented at the 12th IAPR Workshop Document Anal. Syst., Santorini, Greece, Apr. 2016.

[29] O. K. Fasil, S. Manjunath, and V. N. M. Aradhya, ‘‘Word-level script identification from scene images,’’ in Proc. 5th Int. Conf. Frontiers Intell. Comput., Theory Appl., Mar. 2017, pp. 417–426.

[30] A. Nicolaou, A. D. Bagdanov, L. Gomez-Bigorda, and D. Karatzas, ‘‘Visual script and language identification,’’ presented at the 12th IAPR Workshop Document Anal. Syst., Santorini, Greece, Apr. 2016.