International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-2, July 2019

Telugu Word Image Retrieval using Deep learning Convolutional Neural Networks

Kesana Mohana Lakshmi, Tummala Ranga Babu

Telugu word image retrieval (TWIR) is a still challenging task due to the structure complexity of Telugu word image. An efficient TWIR system can be implemented by a holistic representation of word image that comprises of every possible extracted feature. Further, it is also required to retrieve more relevant word images even there is a noisy query word image. Here, it is proposed an efficient TWIR system that utilizes deep learning convolutional neural networks (DL-CNN) to extract the feature map from the query and database word images. In addition, principal component analysis (PCA) is employed to compute the principal features form the feature map and pairwise hamming distance is considered as a similarity metric to retrieve most relevant Telugu word images from the database. Extensive simulation analysis disclosed that proposed TIWR system obtained a superior performance over conventional TIWR systems in terms of mean average precision (mAP) and mean average recall (mAR).

***Index Terms*: Telugu word image retrieval, deep learning convolutional neural networks, principal component analysis, hamming distance, precision and recall.**

I. INTRODUCTION

Document image retrieval is suitable an attractive field of research with the endless development of attention in having information obtainable in the digital format for effective access, reliable storage and long term preservation. Large number of digital libraries such as Universal Library (UL) [1], Digital Library of India (DLI) [2], and Google books are emerging for archival of multimedia documents. These documents cannot be stored as text always. This makes the search for relevant documents even more challenging. Nowadays, the storage devices are suitable cheaper and imaging devices are becoming progressively popular. This motivates researchers to put efforts on developing efficient techniques to digitize and archive large quantity of multimedia data. The multimedia data includes text, audio, image and video. At this stage, most of the archived materials are printed books, and digital libraries are collection of document images. To be more precise, digitized content is stored as images corresponding to pages in books. These documents are typically available in very large numbers hence manually grouping and filing these documents for making them available easily is very tedious task. However, it is very important that these documents are made accessible to the users who would in fact like to search them with relative ease. Scanned/digital form of documents do not comprise searchable text as it is but comprise words as images which cannot be searched/retrieved by current search engines. Traditional text search is based on matching/comparison of textual description (say in ASCII/UNICODE) in association with a powerful language model. These techniques cannot be castoff to access content at the image level, where text is represented as pixels but not as text. The best way of treating these digital documents is to segment the textual content existing in the leaflets. Once the contents are separated out, a representational scheme (profile feature) can be applied to them to get their representational form, which could be ready to use in a content centered image retrieval system. Additionally, a direct way to access these leaflets is by altering document images to their textual form by knowing text from images. Optical character recognizers (OCRs) are required to obtain the textual content from these documents [3, 4]. Success of OCR based text image retrieval schemes mainly depend on the performance of optical character recognition systems (OCRs) [5]. In literature, the use and application of OCR systems are well demonstrated for many languages in the world [6]. For Latin scripts and some of the Oriental languages, high accurate recognition systems are commercially available for use. However, for Indian language with their own scripts, much attention has not been given for developing robust OCRs that successfully recognize diverse printed text images. Therefore, many alternate approaches are presented by researchers to access content of digital libraries in these languages [7]. The emphasis is on recognition free tactics for repossession of related leaflets from great gatherings of file images. In recent years, retrieval of document images from the information of query word has emerged as a vital research field [8-12]. This is also a successful alternative development for the applications of recognition-based handwritten and printed documents in most of the complex scripts retrieval system. Technically, these approaches retrieve the relevant word images from the database by utilizing the extracted features of query word, which will be extracted based on various methodologies with similarity measurement.

Revised Manuscript Received on 30 July 2019.
***** Correspondence Author
Kesana Mohana Lakshmi*, Research Scholar, Dept. of ECE, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur.
Department of Electronics and Communication Engineering, CMR Technical Campus, Hyderabad, Telangana, India
Tummala Ranga Babu, Dept. of ECE, RVR & JC College of Engineering, Guntur.

Generally, these retrieval systems performances depend on how effective the features are derived from the word images and the similarity measuring approach utilized to match with data base word images.

There are plenty of research works presented in the literature of English word images. Investigation of south Indian languages like Telugu has not been presented in much research papers in the field of word image retrieval. For structuring of bag-of-words (BoW) representation, processed features using the scale-invariant feature transform (SIFT) [13] at intrigue emphases are the most well-known features. Later years, bag of visual words (BoVW) is innovated as a concept for presenting and organizing the word images [14, 15]. Author in [14] addressed an efficient word image retrieval using integrated BoVW with SIFT approach. They considered different kind of languages for query word images and obtained superior efficiency over SIFT approach discussed earlier. In [16], hidden markov model with correlation (HMM-C) is utilized for TWIR system and disclosed the significance of HMM in TWIR system. Recently, speed up robust features (SURF) is incorporated with BoVW to obtain an enhanced performance of TIWR system when compared to existing BoVW and SIFT-BoVW approaches.

A further enhancement of TWIR system is achieved in [18], where the authors employed gray level cooccurrence matrix with iterative partitioned clustering (GLCM-IPC) approach. Additionally, image statistics also computed for more accuracy. However, due to the lack of efficiency, lower accuracy and higher complexity, above mentioned TWIR systems are unable to provide relevant word images with noisy, occlusion and random distorted queries. Most importantly, missing segment Telugu word images have not been considered in earlier works. Recently, an efficient approach for Telugu script recognition and retrieval using transformation-based methodology is proposed in [21], which utilized missing segment, noisy, corrupted and occlusion effected word images as a query input, also deliberated multi conjunct vowel consonant gathered word images for showing the robustness of proposed algorithm. However, extraction of features from the word image plays a significant role in retrieval system, which is quite hard in [19-21] and even in other word image retrieval systems in literature. Thus, a DL-CNN with principal component analysis based pair wise hamming distance (PCA-H) is employed to extract the maximum features from the given word image where the extracted feature map represents the word image more effectively and assists for retrieving more relevant word images even with larger databases. Major contributions of proposed TWIR system is described as mentioned below:

- Novel use of deep learning convolutional neural networks (DL-CNNs) for feature extraction of Telugu word images. In addition, PCA is employed for extracting the principal components from the extracted features for further enhancement of document image retrieval system.
- Fully new framework on TWIR process is offered by utilizing CNN approach and various sort of Telugu word images like missing segment, occlusion affected, noisy and random distortions.

The rest of the paper is organized as follows: Section II provides proposed framework. Section III describes the experimental analysis and discussion with conventional TWIR systems provided in the literature. At last, conclusions are specified in section IV followed by references.

II. PROPOSED METHODOLOGY

This section describes proposed methodology which employs DL-CNN for TWIR system. Working of CNN can be explained as follows: A 2-D convolutional layer relates sliding filters to the input. The layer convolves the input by moving the filters alongside the input vertically and horizontally and calculating the dot product of the weights and the input, and then totaling a bias term. A ReLU layer achieves a threshold process to every element of the input, where any value less than zero is set to zero. A max pooling layer performs down-sampling by isolating the input into rectangular pooling regions and calculating the extreme of each region. A fully connected layer multiplies the input by a weight matrix and then complements a bias vector.

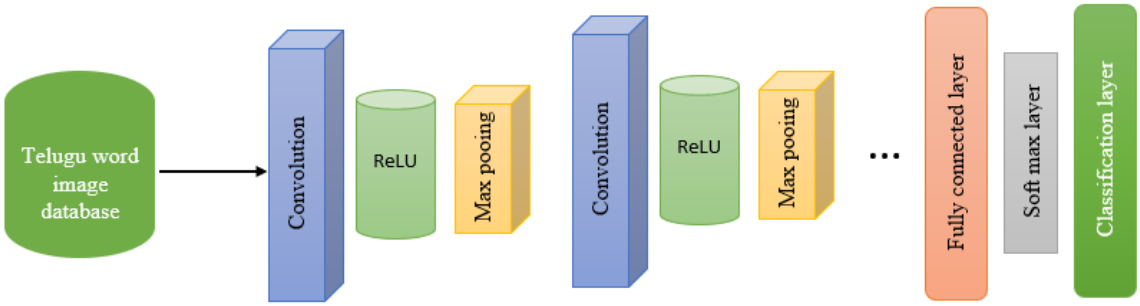


Fig. 1 Architecture of DL-CNN

A. DL-CNN

According to the facts, training and testing of DL-CNN involves in allowing every source image via a succession of convolution layers by a kernel or filter, rectified linear unit (ReLU), max pooling, fully connected layer and utilize SoftMax layer with classification layer to categorize the objects with probabilistic values ranging from $[0,1]$.

Figure 1 discloses the architecture of DL-CNN that is utilized in proposed methodology for TWIR system for enhanced feature representation of word image over conventional retrieval systems.

Convolution layer as depicted in Figure 2 is the primary layer to extract the features from a source image and maintains the relationship between pixels by learning the features of image by employing tiny blocks of source data. It's a mathematical function which considers two inputs like source image $I(x, y, d)$ where x and y denotes the spatial coordinates i.e., number of rows and columns. d is denoted as dimension of an image (here $d = 3$, since the source image is RGB) and a filter or kernel with similar size of input image and can be denoted as $F(k_x, k_y, d)$.

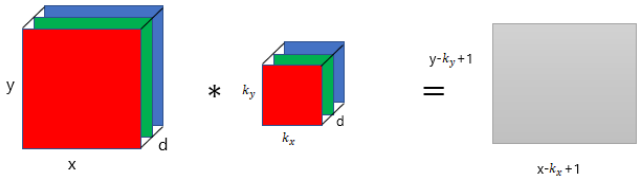
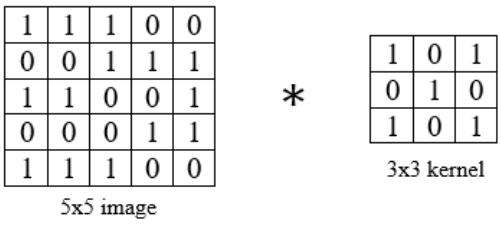
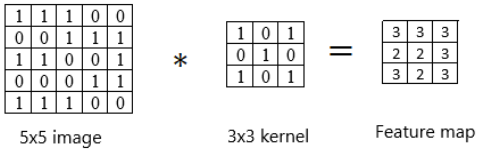


Fig. 2 Representation of convolution layer process

The output obtained from convolution process of input image and filter has a size of $C\left((x - k_x + 1), (y - k_y + 1), 1\right)$, which is referred as feature map. An example of convolution procedure is demonstrated in Figure 3. Let us assume an input image with a size of 5 x 5 and the filter having the size of 3 x 3. The feature map of input image is obtained by multiplying the input image values with the filter values as known in Figure 3.



(a)



(b)

Fig. 3 Example of convolution layer process (a) an image with size 5×5 is convolving with 3×3 kernel (b) Convolved feature map

B. ReLU layer

Networks those utilizes the rectifier operation for the hidden layers are cited as rectified linear unit (ReLU). This ReLU function $G(\cdot)$ is a simple computation that returns the value given as input directly if the value of input is greater than zero else returns zero. This can be represented as mathematically using the function $max(\cdot)$ over the set of 0 and the input x as follows:

$$G(x) = \max\{0, x\}$$

C. Max pooing layer

This layer mitigates the number of parameters when there are larger size images. This can be called as subsampling or down sampling that mitigates the dimensionality of every feature map by preserving the important information. Max pooling considers the maximum element form the rectified feature map.

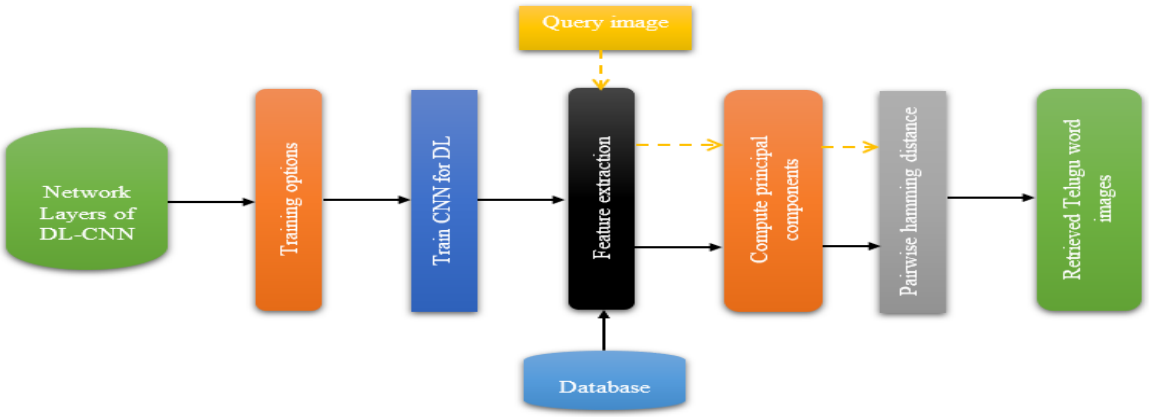


Fig. 4 Proposed DL-CNN-PCA-H for TWIR system

A. Principal component analysis

PCA is an approach of machine learning which is utilized to reduce the dimensionality. It utilizes simple operations of matrices from statistics and linear algebra to compute a projection of source data into the similar count or lesser dimensions. PCA can be thought of a projection approach where data with *m*-columns or features are projected into a subspace by *m* or even lesser columns while preserving the most vital part of source data. Let *I* be a source image matrix with a size of *n* * *m* and results in *J* which is a projection of *I*. The primary step is to compute the value of mean for every column. Next, the values in every column are centered by subtracting the value of mean column. Now, covariance of the centered matrix is computed. At last, compute the eigenvalue decomposition of every covariance matrix, which gives the list of eigenvalues or eigenvectors. These eigenvectors constitute the directions or components for the reduced subspace of *J*, whereas the peak amplitudes for the directions are represented by these eigenvectors. Now, these vectors can be sorted by the eigenvalues in descending order to render a ranking of elements or axes of the new subspace for *I*. Generally, *k* eigenvectors will be selected which are referred principal components or features.

B. Pairwise hamming distance

To assess distances between query word image I_q and retrieved word imagesI_r, a metric necessity be defined and requisite a measurement method to express how the query and retrieved word images are alike (bit per bit). Hence, we want a likeness degree where the distance value will be the number of alike bits in the deliberated Telugu word images. Next table gives similarity truth table for the distance we want to define.

Seeing the nth bit of $I_q$and I_r, and the distance between those two is denoted as $\mathfrak{D}$ then truth table of similarity is as follows:

Table 1. Truth table of similarity measurement using hamming distance metric

$I_q[n]$	$I_r[n]$	$\mathfrak{D}(I_q[n], I_r[n])$	Similarity
0	0	0	relevant
0	1	1	Irrelevant
1	0	1	Irrelevant
1	1	0	Relevant

III. RESULTS AND DISCUSSION

This section designates the simulation analysis and discussion of proposed TWIR system with several test images depicted in Figure 5 and the details of considered Telugu word images are disclosed in Table 2. The testing process and the simulation is done in MATLAB 2018a environment under higher CPU specifications. As shown in Figure 5, different kind of Telugu word images like occlusion affected, missing segment, noisy effected, random distortion and missing segment with random distorted images are considered as a query word images. Assessment of proposed TWIR system using DL-CNN is done by computing mean average precision (mAP) and mean average recall (mAR) and compared with the conventional TWIR systems like SIFT-BoVW [14], HMM-C [16], SURF-BoVW [17], GLCM-IPC [18], HWNET v2 [19] and SDM-NSCT [21]. As discussed earlier, simulation analysis is done with several kind of Telugu word images and obtained enhanced mAP and mAR even when the query word images had a kind of unwanted information which might be introduced automatically while acquiring them or manually by a human or even by a printing machine during the scanning procedure.

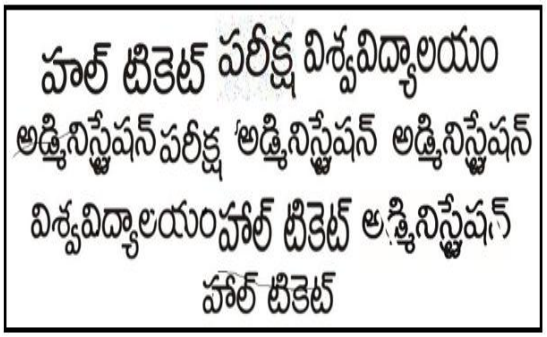


Fig. 5 Test images utilized for TWIR system

Table 2. Book used for experiment

Book	#Pages	#Words
Telugu	700	10116

Figure 6 shows that the retrieved Telugu word images of proposed TWIR system for different kind of query word images using DL-CNN-PCA-H. All the query word images are given in 1st column of Figure 6, an example of Telugu word images with missing segment are disclosed in 1st, 2nd and 3rd row with their retrieved word images. Similarly, a word image with random disturbance and the retrieved Telugu word images using proposed TWIR system is given in 4th row. Next, a noisy query image with its retrieved word images is demonstrated in 5th row. Attained relevant Telugu word images for a query word image with missing segment and random distortion is disclosed in Figure 6(g). Table 3 demonstrates that mAP and mAR values of conventional TWIR system with proposed TWIR system.

Tested query word images	Retrieved word image from database using proposed approach

Fig. 6 Retrieved Telugu word images with DL-CNN (a), (b), (c) Missing segment words. (d) Random disturbance. (e) Occlusion effected. (f) Noisy as a query word. (g) Missing segment with random distortion.

Table 3. Performance comparison of conventional and proposed TWIR systems with mAP and mAR.

Measurement	SIFT + BoVW [14]	HMM-C [16]	SURF+ BoVW [17]	GLCM-IPC [18]	HWNET v2 [19]	SDM-NSCT [21]	Proposed TWIR system
mAP	0.853	0.89	0.731	0.967	0.978	0.998	**0.999**
mAR	0.799	0.823	0.809	0.842	0.92	0.98	**0.999**

IV. CONCLUSION

This article presented an efficient TWIR system using DL-CNN-PCA-H where PCA computed principal features of the feature map obtained from DL-CNN and pairwise hamming distance is computed for similarity measurement. Simulation results disclosed that proposed TWIR system obtained superior performance by retrieving more relevant Telugu word images even with noisy, occlusion, missing segment, random disturbance and some mixed corrupted query word images. Further, the performance evaluation of proposed TWIR system is demonstrated using mAP and mAR and compared with the existing TWIR systems presented in the literature.

REFERENCES

1. Digital library of india. http://dli.iiit.ac.in/.
2. The universal library. http://www.uliborg.
3. D. Bainbriidge, Thompson.J, and Witten.H.I, "Assembling and enriching digital library collections", Conference Proc. on Digital Libraries, USA, pp. 323–334, 2003.
4. G. Nagi, "Twenty years of document image analysis in PAMI", IEEE Trans. on Pattern Analysis and Machine Intelligence, 22 (1), pp. 38-62, 2000.
5. W. Saffady. Introduction to Automation for Librarians. 4th edition, 1999.
6. C.Y.Sueen, Mori.S, "Analysis and recognition of Asianscripts: the SOA", In Proc. of 7th Int. Conf. on DAR, UK, pp. 866-872, 2003.
7. Chaudhuri.BB, Pall.U, "Automatic recognition of Oriya script", Sadhana, 27(1), pp. 23-34, 2002.
8. Y.H.Tayy, M.Khaliid, "Offline cursive HRS based on HMM and neural networks", In Proc. of Int. Symposium on Computational Intelligence in Robotics and Automation, Korea, 2003.
9. C.V. Jawaharr, S.S.Ravi, "A Bilingual OCR for Hindi-Telugu Leaflets and its Applications", In: Proc. of 7th Int. Conf. on DAR, pp. 1-5, Sep. 2003.
10. N.S.Raani, T.Vasuudev, "A GLE Methodology using CM for Printed and Handwritten Document Images", In Proc. of International Conference on Emergent Study in CICA, 3(1), pp. 589-594, 2014.
11. R.Singh, M.Kaur, "OCR for Telugu script using BPBC", Int. Jour. of IT and Knowledge Management, 2(2), pp. 639-643, 2010.
12. S.V.Patgar, T.Vasudevv, "A system for detection of production in photocopy document", 2nd Int. Conf. on CSE, 2015.
13. Lowe.DG, "Distinctive Image Features from SI Key points", Int. Jour. of Computer Vision, 60(2), pp. 91-110, 2004.
14. R. Shekhar and C.V.Jawahaar, "WIR Using Bag of Visual Words", IAPS Int. Workshop on DAS, Gold Cost, QLD, Australia, pp. 297-301, 2012.
15. Yalniz,IZ, R.Manmathaa, "An Efficient Framework for SearchingText in NoisyDocument Images", IAPS Int. Workshop on DAS, Gold Cost, QLD, Australia, pp. 48-52, 2012.
16. Nagasudhha.D, Y.M.Lattha, "Keyword Spotting using HMM in Printed Telugu Documents", In Proc. of Int. Conference on SCOPES, Paralakhemundi, India, pp:1997-2000, Oct. 2016.
17. N.Jayanthi, S.Indhu, "Inscription IR Using Bag-of-Visual Words", IOP Conf. Series: MSE, 225(1), pp.1-8, 2017.
18. K. M. Lakshmi and T. R. Babu, "A New Hybrid Algorithm for Telugu Word Retrieval and Recognition", Int. Jour. of IES, vol. 11, no. 4, pp.117-127, 2018.
19. P. Krishnaann, C.V.Jawahaar, "HWNET v2: An efficient word image representation for handwritten documents", Computer Vision and Pattern Recognition, 2018. **arXiv:1802.06194 [cs.CV]**
20. K. Dutta, P.Krishnaann, M. Mathew, "Improving CNN-RNN hybrid networks for handwritten recognition", In: International Conference on Frontiers and Handwriting Recognition, Niagara Falls, NY, USA, Dec. 2018.
21. K. M. Lakshmi and T. R. Babu, "Robust algorithm for Telugu word image retrieval and recognition", Journal of Mechanics of Continua and Mathematical Sciences, vol. 14, no. 1, Feb. 2019.