
Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2024 SCPE. Volume 25, Issues 5, pp. 3741–3751, DOI 10.12694/scpe.v25i5.3176

RETRIEVAL OF TELUGU WORD FROM HAND WRITTEN TEXT USING DENSENET-CNN

RAJASEKHAR BODDU *AND EDARA SREENIVASA REDDY †

Abstract. The recognition of telugu hand written text is been one of the problems in many applications. To overcome the problem a deep learning technique is proposed in this work i.e. a Dense convolutional neural network (DCNN) model. A telugu dataset which is taken form IIIT-HW-Telugu is utilized to perform the proposed model. In this paper a four stage telugu word retrieval is performed, initially thinning of image is performed using morphological operation, secondly Densenet-CNN is applied for thinning image, thirdly perform OCR based image segmentation, finally two models like HARRIS and BRISK features to extract the features and evaluate information from the given input HWT images. The parameters evaluated are hamming distance, PSNR, MSE, Noise Sensitivity and rate of thinning. The proposed model outperformed well compared to other methods. The PSNR obtained using proposed model is 54.74, hamming distance is 1.2.

Key words: OCR, Morphological operations, Hilditch transform, CNN, DenseNet

1. Introduction. The never-ending desire to liberate information to flow in a digital format for easier access, dependence on archiving and preserving for longer term makes hand written text based word retrieval a highly intriguing subject of research. A large variety of digital libraries are emerging for the archiving of multimedia documents, including Universal Library (UL), Digital Library of India (DLI), and Google Books. These documents cannot always be saved as text. This increases the difficulty of finding pertinent papers. The cost of storage devices has decreased, and imaging devices are rising in popularity. This encourages scientists to work hard on creating effective methods for digitising and archiving massive amounts of multimedia material. Text, audio, picture, and video are all included in the multimedia data. Most of the items that have been archived so far are in books printed, while digital libraries are currently collections of document pictures. More specifically, digital material is saved as pictures that match to book pages.

Daily existence requires the use of images [1]. Every day, a large amount of data is created in the form of photographs by technological devices. The two main categories of image retrieval [2] are approaches based on text and second one is based in content. Each strategy has distinctive qualities [3] that may be applied to a variety of applications depending on the circumstance. Search by image content is discussed by author in [4] and is termed as content-based image retrieval (CBIR). With this method, the search examines the image's visual content rather than the keyword descriptions linked to it. Text-based image retrieval (TBIR) [5] relies only on keywords based on text and input to be descriptors, with text keywords also employed in the index pictures. The matched photos are fetched from the image repository by comparing the index images keywords with the supplied keywords. This strategy has the benefits of being simple to implement, quick retrieval, and web picture search. It is difficult to manually search through a vast collection of photographs for any image.

The author in [6] discussed about the study of computer vision and machine learning, where machine learning helps in developing the automated system of identification of scene from the text. This recognition of text in natural settings is a big challenge. English has been the primary language of research in the field of text categorization and domain identification. Regional languages, particularly Indian languages, have had far less of an impact. Telugu is a member of the Dravidian language family and is one of language which is older and traditional of Indians. Telugu is the sixteenth most spoken language in the world, with 93 million native

*Research Scholar, Department of Computer Science and Engineering, College of Engineering and Technology, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. (rajsekhar.se15@gmail.com).
†Professor, Department of Computer Science and Engineering, College of Engineering and Technology, Acharya Nagarjuna University, Guntur, Andhra Pradesh, India.

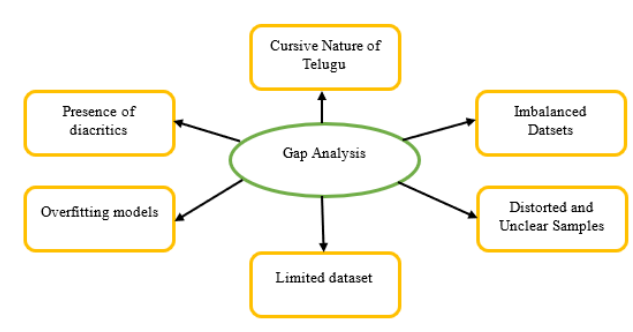

Fig. 1.1: Gap Analysis of Telugu Text Recognition [21]

speakers, according to the Ethnologic list. The gap analysis for any language is same as shown in figure 1.1.

It became a difficult process to retrieve telugu words from handwritten writing. Compared to other languages, telugu has a particular set of writing curves and strokes for its handwritten text. One of the cutting-edge methods for addressing issues in the sector of concern is deep learning. With the use of massive quantities of data and the Deep Learning technology, the nodes of one layer are connected to those of adjacent layers. The network is assumed to be more complex due to the number of tiers. Since deep learning systems handle massive volumes of data and perform challenging mathematical operations, they demand strong hardware. In this paper the utilization of DenseNet CNN is considered for effective retravel of telugu word from hand written text. Further in section 2 the discussion is about different techniques used previously. In section 3, the suggested framework is implemented and in section 4 the results obtained using proposed model is discussed.

2. Related Work. For the characterisation or acknowledgement of word pictures for different dialects, several writings are accounted for. When compared to the typical scanned texts in English writing, acknowledgment of terms from Telugu writings has not been studied as extensively. A part of the approaches mentioned are briefly discussed. The most popular highlights for creating bag-of-words (BoW) representation are those processed at interest focuses in scale-invariant feature transform (SIFT) [7] highlights. A histogram of the visual words serves as the word image's visual representation. "Data is lost when the highlights are quantized to a visual word. This is frequently believed to exhibit some measure of power (or invariance). In fact, there is still no consensus on how to select the vocabulary's range and retention techniques. Given a language, creating a BoW representation requires two key steps one is Coding and the other is pooling. With the use of the vocabulary words' histograms of repeated incidents, reports are properly expressed. The categorization and recovery of records are then carried out using these histograms.

Spatial pyramid matching (SPM) [8], which divides pictures into vertical and flat bearings, provides spatial request in common scene images. Word pictures are divided vertically into three portions and then recorded to provide order in representation. In [9], word pictures are shown as profile highlights, and Euclidean separation is unintentionally employed to imply comparison. Dynamic time wrapping (DTW) is used to coordinate word pictures in order to account for the variance in word image lengths. However, the focus of this study is to deduce an enforcement conspiracy using a back-end file structure. Due to this, DTW-based systems are not rational. Versatility in report recovery has also recently received significant attention. A list of 10 million pages using locally likely arrangement hashing (LLAH) was published in [10]. Recent efforts to recover strong records use the visual Bag of Words (BoVW) to represent and organise word pictures. One may quickly find significant papers from a million documents using BoVW representation and a reversed ordering scheme [11]. For the picture representation, feature points will be quantized, and a versatile representation is created by proposing a "vocabulary" over an element space. Ideally, code is generated from raw descriptors using vector quantization (VQ). The difficult issue of translating code words from the vocabulary to the feature vectors of an image is one limitation of the code-book technique. The challenging assignment presents the two problems of codeword credibility and vulnerability. The problem of selecting the proper codeword from at least two

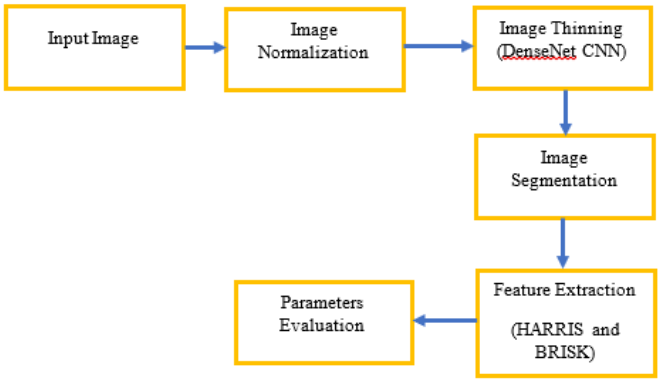

Fig. 3.1: Block Diagram of Proposed Work

significant candidates is known as codeword vulnerability. The VQ technique disregards the significance of several candidates and selects the most effective visual word. The problem of picking a codeword without a plausible hope in the lexicon is known as "codeword believability." The codebook method assigns the best-fitting codeword, although it is not a true representation. Authors in [12] proposed a soft assignment coding approach to overcome this restriction, in which each visual word is assigned a local characteristic based on its location.

In [13] author represented a spotting system of a Telugu word with improved performance over traditional system. It is based on correlation and hidden markov model (HMM) technology. In order to outperform BoVW and SIFT + BoVW, an algorithm is developed based on sped-up robust features (SURF) with the aid of BoVW. However, the word picture retrieval techniques described in the literature have some drawbacks, including inefficiency, increased complexity, and decreased precision with big data bases.

Contrarily, people in the south Telugu, an Indian alphabet, is composed of several elements, making the use of high-level feature extraction algorithms more difficult. On Indian languages, several techniques for domain identification and text classification have been developed, however only a small number of these studies have been reported on Telugu. This section provides an overview of a few strategies and approaches for text classification and domain identification. Automated text classification with a focus on Telugu is been developed in recent times. In his research, 800 Telugu news items were classified using supervised classification with the Naive Bayes classifier. KNN, Naive Bayes, and decision tree classifiers as text mining approaches to represent and categorise papers written in Indian. Telugu text documents can be categorised using the language-dependent and independent models suggested. Telugu texts were classified using a model for document organisation and categorization of texts using the word frequency ontology. The robustness of LSTM and CNN are combined in an attention-based multichannel CNN for text classification. In this network, CNN tracks word relationships while Bi-LSTM records word history and future information [14]. The author in [15] suggested a novel heuristic advanced neural network based telugu text categorization model (NHANNTCM) for extraction of telugu word and achieved an accuracy of 99%.

3. Methodology. In this paper thinning of the input image is the main concept. The thinning is performed using DenseNet CNN and followed by OCR based segmentation of letters and finally extraction of features using Harris and Brisk. The major goal of this study is to take use of the DL-CNN's robust performance in order to extract useful features with the least amount of time and effort possible. The convolutional phase known as CNN extracts features from pictures by acting as a visual descriptor. Each image is altered by the application of a series of filters, resulting in new image types known as convolution maps. The proposed model is shown in figure 3.1.

3.1. Input Dataset. Handwriting recognition (HWR) in Indic scripts is a challenging problem due to the inherent subtleties in the scripts, cursive nature of the handwriting and similar shape of the characters. Lack

Fig. 3.2: Example of input words

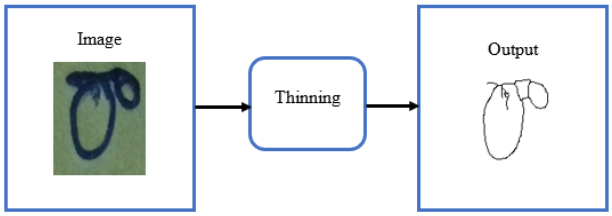

Fig. 3.3: Concept of Thinning

of publicly available handwriting datasets in Indic scripts has affected the development of handwritten word recognizers. In order to help resolve this problem, 2 handwritten word datasets: IIIT-HW-Dev, a Devanagari dataset and IIIT-HW-Telugu, a Telugu dataset [16]. In this paper, Telugu Dataset (IIIT-HW-Telugu) with a file size of 3.7 GB is considered to evaluate the proposed model. The samples of images considered as input is shown in figure 3.2.

3.2. Image Normalization. To extract various features from a picture on the same structure for image normalisation, all randomly sized images are downsized into the same size images. Here, using a bilinear standard transformation, photos of arbitrary sizes are normalised to be 100x200 size and the aspect ratio of image should not get disturbed. The numerous distortions, including missing segments with distortion, distortion caused randomly, effects of noise, and the segments which are missing in query terms, are also calculated, and analysed using this normalisation procedure.

3.3. Thinning of Image. Thinning is the technique of removing unused pixels from an image in order to recover the skeletons. It is also known as the Skeletonization process. Black foreground pixels are removed repeatedly, layer by layer, in this morphological procedure until a one-pixel-wide skeleton is reached. It entails reducing something to its tiniest size. Typically, binary pictures made up of black (foreground) and white (background) pixels are subjected to skeletonization. It accepts a binary picture as input and outputs another binary image, as seen in Figure 3.3.

The use of a deep convolutional neural network to thin an input picture is covered here. When compared to other algorithms, DCNN-based image thinning produces accurate Skelton estimates of the input picture. The deep network utilized in this work is densenet.

3.4. DenseNet CNN. Following is an explanation of how CNN operates: using the two-dimensional layer of convolution given input is paired up with sliding filter. The convolution of layers for the input is computed using the dot products of weights and input, in this process the set of filters are moved in vertical direction and horizontal direction towards the input. Threshold process is done in the ReLU layer by converting the values to zero which are lower than zero. Later down sampling is performed in the max pooling layer by identifying the maximum level of every zone by splitting the input into a rectangle pooling region. Bias vector is added to the fully connected layer, before performing this addition the input is multiplied by a weight matrix. Figure 3.4 depicts the overall architecture of deep CNN.

Here, we will provide a summary of the DenseNet design for convolutional networks, which stands for densely linked convolutional networks. The issue that the convolutional neural network is seeking to address with the density of the design is to deepen it. Dense nets are networks of convolution with plenty of connections.

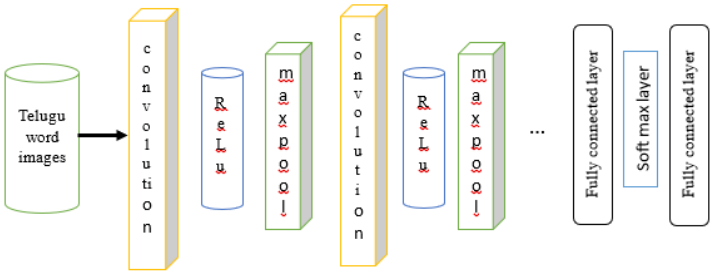

Fig. 3.4: Architecture of DCNN

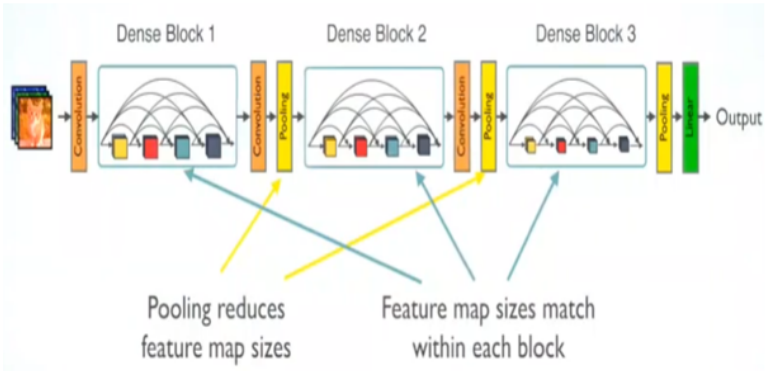

Fig. 3.5: DenseNet Architecture [16]

It is quite like a ResNet, with a few key differences. While ResNet employs an additive technique and accepts a previous output as an input for a subsequent layer, DenseNet utilizes every previous output as a source for a new layer. Figure 3.5 conveys the detailed structure of DCNN.

The drawback of this network is that it gets sort of unsustainable as we go further into it. For example, if the second layer is set for moving towards third layer, then the third layers need to be sourced with the second layer and the other layers from the initial state. In this network dense blocks are created for which different filters are created at every block but the size of feature map are well in constant for every block internally. The layer present in dense network is transition layer, these layers are handled by down sampling. This down sampling is performed by applying normalization of batch, one to one convolution and two to two layers of pooling.

The thinning of input image is performed deeply by carrying every layer of the telugu hand written word to the next layer until a better output is achieved. Figure 3.6 the densenet performance helps in thinning the image with more accurate output based on the working performance of densenet.

3.5. Image Segmentation. In this section, we will go over the segmentation techniques, which are yet another crucial stage of the OCR system. Simply divided into smaller segments for subsequent processing, segmentation is the act of taking a whole picture. Using word level segmentation, a picture is segmented. The input is divided into separate letters, making it easier to recognise the word.

We are given an image with a single line made up of a string of letters at this level of segmentation. As seen in Figure 3.7, the goal of word level segmentation (WLS) is to separate the picture into its component letters.

3.6. Feature Extraction. Feature extraction forms an important part of in retrieval of words from hand written text. In many applications, the speed of feature detection in a picture is critical. To compute the

Layers	Output Size	DenseNet-121	DenseNet-169	DenseNet-201	DenseNet-264
Convolution	112 × 112	7 × 7 conv, stride 2			
Pooling	56 × 56	3 × 3 max pool, stride 2			
Dense Block (1)	56 × 56	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6	[1 × 1 conv, 3 × 3 conv] × 6
Transition Layer (1)	56 × 56	1 × 1 conv			
	28 × 28	2 × 2 average pool, stride 2			
Dense Block (2)	28 × 28	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12	[1 × 1 conv, 3 × 3 conv] × 12
Transition Layer (2)	28 × 28	1 × 1 conv			
	14 × 14	2 × 2 average pool, stride 2			
Dense Block (3)	14 × 14	[1 × 1 conv, 3 × 3 conv] × 24	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48	[1 × 1 conv, 3 × 3 conv] × 64
Transition Layer (3)	14 × 14	1 × 1 conv			
	7 × 7	2 × 2 average pool, stride 2			
Dense Block (4)	7 × 7	[1 × 1 conv, 3 × 3 conv] × 16	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 32	[1 × 1 conv, 3 × 3 conv] × 48
Classification Layer	1 × 1	7 × 7 global average pool			
		1000D fully-connected, softmax			

Fig. 3.6: DenseNet Description [20]

Fig. 3.7: OCR Image Segmentation

correspondence between numerous perspectives effectively and accurately, the detected feature points must be represented independently. Fast feature recognition, description, and matching are necessary for real-time processing of the pictures. The points used to characterise the images must meet two crucial requirements in order to achieve better feature matching of image pairs: first, the feature points of the same strokes in various perspectives, viewpoints, or lighting conditions must be the same; second, the points must have enough information to match with one another. The finest characteristics for matching are corners. The most crucial aspect of a corner is that if one exists in a picture, the surrounding area will abruptly alter in intensity.

The information of pixels which is present locally are explained using local feature descriptors. These local features which are need to be evaluated meet various criteria like blurriness, presence of noise, translation invariant, rotation, scale and transformation based on affline. An effective feature detection operator that has seen widespread application is the Harris corner detector and Brisk corner detection. The rotation invariant Harris corner detector has sufficient data for feature matching.

Due to the Harris corner detector's great invariance to rotation, scale, illumination fluctuation, and noise in image it is a well-liked interest point detector. The local autocorrelation function of a signal serves as the foundation for the Harris corner detector, which monitors local variations in the signal with patches that have been slightly displaced in various directions. The Harris approach looks at the intensity which is average and to be directional, to locate the corners in the input picture. The approach of detecting corners mathematical formulation essentially determines the intensity difference in every direction using a displacement of (u, v).

For the pixel with displacement (u,v) the grey intensity is termed as I(x,y). Here the variation of the pixel

Table 4.1: Requirement of Design Environment

Description	**Requirement**
RAM	8GB
Processor	Intel i7
Matlab version	2021a
Image format	JPEG

that is gray is (x,y) with a shift range of (u,v) is given by equation (3.1).

$$H(u,v) = \sum_{x,y} w_f(x,y)\left[I(x+u, y+v) - I(x,y)\right]^2 \tag{3.1}$$

The term $w_f(x,y)$ denotes the windowing function, the shifted intensity value which is termed as $I(x+u, y+v)$ and the intensity value is termed as $I(x,y)$.

Key point descriptor with scale Scale-spacing and binary description are both handled by the BRISK approach [17]. In the picture pyramid's octave layers, key points are found. Quadratic function fitting is used to translate the coordinates and scale of each key point into representation of a continuous domain. The BRISK descriptor is generated as a binary string in two steps after the BRISK characteristics have been detected. The first stage aids in the creation of a rotation-invariant description by estimating the key points' orientation. To effectively and quickly construct a description that captures regional attributes, the second stage utilises rigorous brightness comparisons. The BRISK descriptor uses a concentric circle sampling method to specify N locations.

The smoothing of intensity at every point s_{pi} is done performed using a gaussian function for preventing of effects like aliasing. The sample points with N number are paired into (s_{pi}, s_{pj}) and are bifurcated into two degrees of classes: one is short pair in which the distance condition should be $(s_{pi}, s_{pj}) < T_{max}$ and the other one is long pair with a distance condition of $(s_{pi}, s_{pj}) > T_{min}$. These two pairs perform individual action like estimation of rotation using the short pair and building of descriptor after correction of rotation is performed using the long pair. The computation of local gradients of BRISK descriptor is given by

$$\nabla(s_{pi}, s_{pj}) = (s_{pj} - s_{pi})\frac{I(s_{pj} - \sigma_j) - I(s_{pi}, \sigma_i)}{\|s_{pj} - s_{pi}\|^2} \tag{3.2}$$

The local gradient is termed as $\nabla(s_{pi}, s_{pj})$ which is the sampled pair and the intensity that is smoothed at x at scaling factor σ. In the average gradients of x and y direction the rotation angle θ is calculated. To get the descriptor that is rotation-invariant, the short pairs are rotated by an angle of $-\theta$. The binary descriptor, which serves as a description for each keypoint, is an encoded binary string.

Algorithm
Step1. Loading the image from the dataset
Step2. Resize all the images and reduce noise
Step3. Removing unused pixels from an image in order to recover the skeletons.
Step4. Using DCNN for extraction of features
Step5. Divide the letters for the given input word
Step6. Evaluate the BRISK features and HARRIS corner points
Step7. Calculate all the parameters like hamming distance between the pixels of the word, MSE, PSNR etc..

4. Experimental Results. The analysis of suggested model is detailed with the help of matlab simulation. The entire simulation is performed using the image of hand written telugu words. The results shown below gives the effectiveness of the proposed model. The consideration for performing simulation is shown in table 4.1 and some of the assumption in simulation environment is shown in table 4.2.

Table 4.2: Assumption in Simulation Environment

Assumption	**Description**
Stability in performance	The evaluation does not undergo any adverse changes that effect the experimental findings
Range of Consistency	The environment of simulation remains consistency and do the parameters
Constant parameters	The parameters used for evaluation are same for all set of input images
No Interference from external source	As there is not external interference the simulation results will be accurate

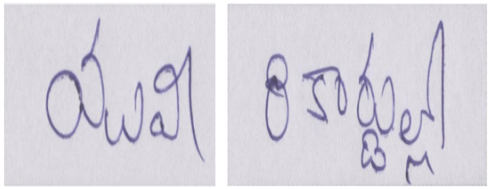

Fig. 4.1: Input Image

Fig. 4.2: Thinning Image

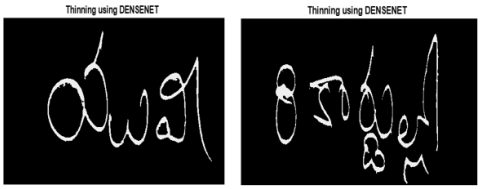

Fig. 4.3: DenseNet based thinning

The input is normalised and sent for next stage of processing that is thinning can be considered from Figure 4.1. The tinning image which is obtained using different techniques is shown in Fig. 4.2 and Fig. 4.3. In this process the front end of the image is highly viewed for achieving better set of features in next stage.

Every letter in the word is been segmented using OCR word segmentation model and the results achieved is shown in fig 4.4.

Harris features and Brisk features are been extracted and is shown in fig 4.5. The features are extracted for the image which is thinned using DenseNet CNN. The duration of telugu data retrieval is less when compared to other extraction of thinning image. The performance will be improved when the features are extracted for DCNN thinned image.

Certain parameters are considered for showing the performance of the suggested model with other existing techniques. The parameters are measure of connectivity, MSE, PSNR, Rate of thinning, RMSE, time of execution, Noise sensitivity and hamming distance. These parameter values are shown in table 4.3.

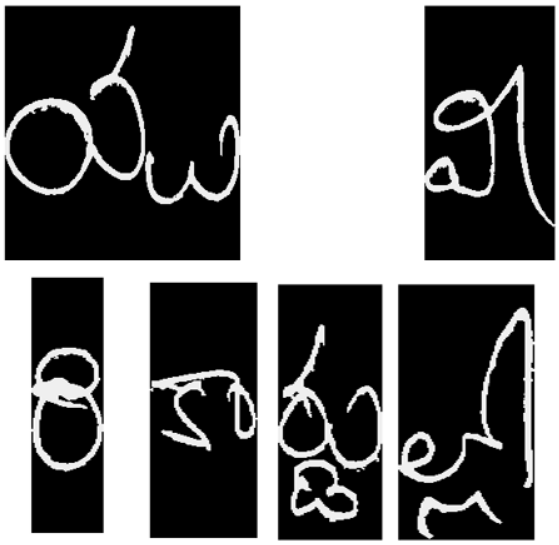

Fig. 4.4: Segmentation of DCNN thinned image

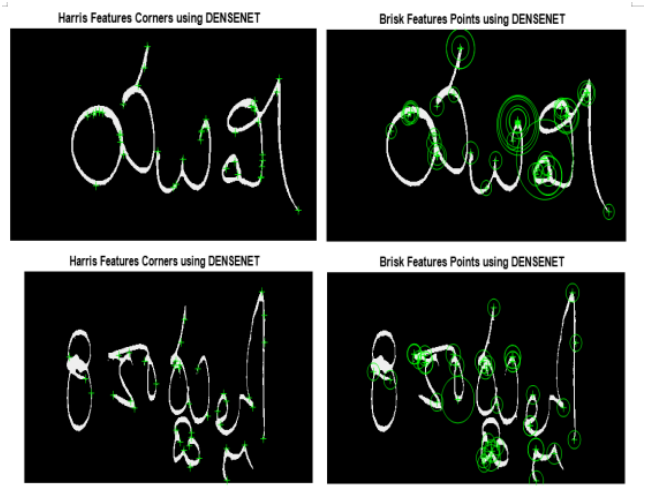

Fig. 4.5: Outputs of feature extraction

One of the crucial metrics for assessing the effectiveness of the employed strategies is PSNR. The PSNR, which affects pixel quality, is the ratio of the highest pixel value to the noise (MSE). The error's value, which is expressed on a logarithmic decibel scale, decreases with increasing PSNR. Figure 4.6 shows the contrast in PSNR.

5. Conclusion. In this study, an efficient model for Telugu word recognition is presented through the testing of several neural network designs. The provided input is first normalised and thinned before being submitted to the deep convolutional neural network model to extract the feature maps. Using DenseNet convolutional neural networks, a model for Telugu text extraction and identification is built in this article. Additional corner characteristics are retrieved using the Harris and Brisk techniques. The simulation research

Table 4.3: Performance measures using different techniques

Measure Evaluted	**Hilditch Algorithm**	**Morphological operations**	**RNCNN-BRHA [18]**	**Proposed HWTR-DCNN**
Measure of Connectivity	4.0	4.0	4.0	**4.01**
Rate of Thinning (pixels)	1.0	1.0	1.0	**1.0**
MSE	0.022	0.025	0.0007	**0.00018**
PSNR	16.49	15.95	51.84	**54.74**
RMSE	0.149	0.159	0.0025	**0.00135**
Time for executing (Sec)	2.19	0.608	0.102	**0.094**
Noise Sensitivity	0.63	0.388	0.50	**1**
Hamming Distance (pixels)	7.15	6.20	2.71	**1.20**

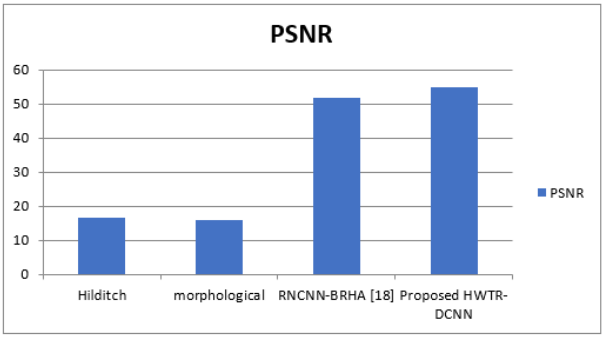

Fig. 4.6: PSNR Comparison

showed that in terms of PSNR, MSE, Noise sensitivity, and execution time, the new technique outperformed the traditional retrieval system. Additionally, the suggested HWTR-DCNN system's performance evaluation is shown using mAP and mAR and is contrasted with the current systems.

REFERENCES

[1] Li, Ang, et al. "Generating holistic 3d scene abstractions for text-based image retrieval." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017.
[2] Unar, Salahuddin, et al. "Detected text-based image retrieval approach for textual images." IET Image Processing 13.3 (2019): 515-521.
[3] MK, Yanti Idaya Aspura, and Shahrul Azman Mohd Noah. "Semantic text-based image retrieval with multi-modality ontology and DBpedia." The Electronic Library (2017).
[4] Zeng, Mengqi, et al. "CATIRI: An efficient method for content-and-text based image retrieval." Journal of Computer Science and Technology 34.2 (2019): 287-304.
[5] Estrela, Vania Vieira, and Albany E. Herrmann. "Content-based image retrieval (CBIR) in remote clinical diagnosis and healthcare." Encyclopedia of E-Health and Telemedicine. IGI Global, 2016. 495-520.
[6] Harmandeep Kaur, and Munish Kumar, "A Comprehensive Survey on Word Recognition for Non-Indic And Indic Scripts," Pattern Anal Applic, vol. 21, pp. 897–929, 2018. Crossref, https://doi.org/10.1007/s10044-018-0731-2
[7] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision, Vol. 60, No. 2, pp. 91-110, 2004.
[8] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", In: Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition, 2006.
[9] T. M. Rath and R. Manmatha, "Word spotting for historical documents", International Journal of Document Analysis and Research, Vol. 9, No. 2-4, pp.139-152, 2007.
[10] K. Takeda, K. Kise, and M. Iwamura, "Real-time document image retrieval for a 10 Million pages database with a memory efficient and stability improved LLAH", In: Proc. of the International Conf. on Document Analysis and Recognition, pp. 1054-1058, 2011.

[11] R. Shekhar and C. V. Jawahar, "Word Image Retrieval Using Bag of Visual Words", In: Proc. of the International Workshop on Document Analysis Systems, pp. 297-301, 2012.
[12] J. van Gemert, J.-M. Geusebroek, C. J. Veenman, and A. W. M. Smeulders, "Kernel Codebooks for Scene Categorization", In: Proc. of European Conf. on Computer Vision, pp. 696-709, 2008.
[13] D. Nagasudha and Y. M. Latha, "Keyword Spotting using HMM in Printed Telugu Documents", In: Proc. of International Conf. on Signal Processing, Communication, Power and Embedded Systems, pp. 1997-2000, 2016.
[14] Zhenyu Liu, Haiwei Huang, Chaohong Lu, and Shengfei Lyu. 2020. Multichannel cnn with attention for text classification. ArXiv, abs/2006.16174.
[15] Rajasekhar Boddu, Edara Sreenivasa Reddy (2023). Novel Heuristic Recurrent Neural Network Framework to Handle Automatic Telugu Text Categorization from Handwritten Text Image. International journal of recent and innovation trends in computing and communication, vol.11, No.4, pp.296-305.
[16] http://cvit.iiit.ac.in/research/projects/cvit-projects/indic-hw-data
[17] Leutenegger S, Chli. M, Siegwart RY (2011) Brisk: Binary robust invariant scalable keypoints. In: Proceedings of the 2011 International Conference on Computer Vision, IEEE Computer Society, Washington, DC, USA, ICCV '11, pp 2548–2555.
[18] Boddu, R., Reddy, E.S. (2023). Fusion of RNCNN-BRHA for recognition of telugu word from handwritten text. Revue d'Intelligence Artificielle, Vol. 37, No. 1, pp. 215-221.

Edited by: Anil Kumar Budati
Special issue on: Soft Computing and Artificial Intelligence for wire/wireless Human-Machine Interface
Received: Jan 12, 2024
Accepted: Mar 20, 2024