A Deep Neural Network-Based Approach for Extracting Textual Images from Deteriorate Images

Binay Kumar Pandey[1,*], Digvijay Pandey[2], Subodh Wariya[3], Gaurav Agarwal[4]

[1]Department of Information Technology, College of Technology, Govind Ballabh Pant University of Agriculture and Technology Pantnagar, Uttarakhand, India.

[2]Department of Electronics Engineering, Institute of Engineering and Technology, Dr. A.P.J. Abdul Kalam Technical University, Lucknow, India.

[3]Department of Electronics Engineering, Institute of Engineering and Technology, Lucknow, India

[4]Invertis University Bareilly, U.P India.

Abstract

INTRODUCTION: The quantity of audio and visual data is increasing exponentially due to the internet's rapid growth. The digital information in images and videos could be used for fully automated captions, indexing, and image structuring. The online image and video data system has seen a significant increase. In such a dataset, images and videos must be retrieved, explored, as well as inspected.

OBJECTIVES: Text extraction is crucial for locating critical as well as important data. Disturbance is indeed a critical factor that affects image quality, and this is primarily generated during image acquisition and communication operations. An image can be contaminated by a variety of noise-type disturbances. A text in the complex image includes a variety of information which is used to recognise textual as well as non-textual particulars. The particulars in the complicated corrupted images have been considered important for individuals seeing the entire issue. However, text in complicated degraded images exhibits a rapidly changing form in an unconstrained circumstance, making textual data recognition complicated

METHODS: The naïve bayes algorithm is a weighted reading technique is used to generate the correct text data from the complicated image regions. Usually, images hold some disturbance as a result of the fact that filtration is proposed during the early pre-processing step. To restore the image's quality, the input image is processed employing gradient and contrast image methods. Following that, the contrast of the source images would be enhanced using an adaptive image map. Stroke width transform, Gabor transform, and weighted naïve bayes classifier methodologies have been used in complicated degraded images to segment, features extraction, and detect textual and non-textual elements.

RESULTS: Finally, to identify categorised textual data, the confluence of deep neural networks and particle swarm optimization is being used. The dataset IIIT5K is used for the development portion, and also the performance of the suggested methodology is assessed by utilizing parameters like as accuracy, recall, precision, and F1 score. It performs well enough for record collections such as articles, even when significantly distorted, and is thus suitable for creating library information system databases

CONCLUSION: A combination of deep neural network and particle swarm optimization is being used to recognise classified text. The dataset IIIT5K is used for the development portion, and while high performance is achieved with parameters such as accuracy, recall, precision, and F1 score, characters may occasionally deviate. Alternatively, the same character is frequently extracted [3] multiple times, which may result in incorrect textual data being extracted from natural images. As a result, an efficient technique for avoiding such flaws in the text retrieval process must be implemented in the near future.

*Corresponding author. Email: Binaydece@gmail.com

1. Introduction

The intricate deteriorated image document is a widely accessible and useful media that contains crucial, helpful data that need to transfer on wireless network [55]. These data are composed of pixels, of which important information is extracted from complex images according to the requirements of the computer vision [50]. Complex videos and complex image text messages produce quality, just like text messages in images are used in many complex image learning [43] and indulging implementation processes, including language translators, book digitization, video or image recovery [44]. Considerable devotion has been paid recently to the use of pre-programmed text detection and text recognition [33] over wireless network [54]. In the past, a lot of research was carried out to remove text from transmitted complex scenes images over communication network, and this process of text acquisition is one of the most important parts of OCR [30]. After text detection and further binarization OCR [26] is used to recognize text from images [16].

The visual images provide accurate and appropriate details for oblivious direction-finding, image perceptive, and retrieval methods, respectively [13]. This complex image frequently incorporates a variety of fonts and other properties [2]. The primary focus of the complicated corrupted images could include distinct-designed characters, information exposed in digital signpost information displayed on a monitor. This is a very common task for traditional OCR: identifying textual information with a different appearance. The texts are sprinkled throughout this complex degraded image, and the preceding information about their position is not presented. The input documents from the camera recognize the line spacing number, character, but the complex image text does not contain any formatting rules, so it is not possible to directly apply the segmentation approach to complex images [22]. The process of detecting text in complex degraded images consists of two major steps: classification of textual image and non-textual image and character recognition [28].

A primary goal of textual information extraction [39] from images would be generally used for directing visitors, locating vehicle locations, visually impaired people, and so on. Throughout this approach, an effective DNN is incorporated with the purpose of fetching of the text from tarnished image [29]. In advance to that, the distorted frame of image must be evaluated to determine whether or not the usable image contains any relevant information. To achieve this, a weighted [11] have based classifier-based method is used before using deep learning-based [27] recognition of characters. weighted naive based classifier distinguishes between textual and non-textual images [7]. The error caused by reassessment mostly during the text classification process can be decreased by further using an optimization algorithm. Finally, an image containing the text is fed into the DNN algorithm for text extraction. Furthermore, the load parameter available in DNN [1] during the text extraction method reduces the likelihood of reliability in DNN. To avoid such a situation, a hybrid approach combining particle swarm optimization and DNN has been used, that further results in an optimal weight parameter for DNN.

The following is how this paper is structured. The literature review is illustrated in section 2. Section 3 is concerned with problem statement and motivation. Section 4 deliberates the suggested methodology. Finally, section 5 summaries the proposed methodology's results.

2. Literature Review

This section introduces a numerous approach for textual data classification and character identification have been explored, and the effectiveness of all these approaches has also been evaluated.

In [20] the detection of the bowed and multifaceted textual data in a unified type from complex degraded images frame via evolving a new mask region-based convolutional neural network-based text detection method [45]. However, according to [8] a new method for data recognition and detection has been a connected component- focused on method that made use of the maximally stable extreme regions. The multiple blurs produced by motion and defocus makes the text detection process a challenging one.

In this [47], a method for text identification process in a distorted or non-distorted image is discussed and the contrast variants experienced in nearby pixels were identified in this method to evaluate the blur degree, moreover, the low pass filter was used for de-blurring. Mostly this approach gives pixels under consideration for de-blurring images. The process of detecting the scene text from videos attains high value in several information removal hinged audiovisual applications like video-frame recovery and investigation.

This [45], gives text recognition methods aimed at frames of videos [48]. The public scene text video was included in this method which outperforms the other existing methods.

According to [24], texture data was retrieved and analyzed from complex degraded images using an improved algorithm. To begin, the Discrete Wavelet Transform (DWT) was used to detect edges in images. Following that, the textual regions were located using the Ada Boost classification model and connected component clustering.

In [40] morphological reconstruction is introduced, which is based on a technique known as geodesic transform, which emphasizes artifacts in the image's center while erasing light and dark constructions that are problematic near the image's boundary. This binarization technique has been found to be far superior to other text binarization techniques.

This [51] discussed the utilization information of edge for identifying textual blocks from grayscale images. Its primary goal is to detect text in noisy images and distinguish

it from graphical images. In this case, an algorithm is created to extract features from various objects and then classify those feature points to identify textual regions. Directionally placed text blocks can be easily obtained by using methods such as line of approximation and layout categorization. In the final step a feature dependent, connected component is merged with alike textual parts exist inside the bounding rectangles. The methods anticipated here yield promising results, demonstrating the method's effectiveness.

In [41], a binarization technique for colour images is discussed, and it is discovered that the traditional method, which is based on thresholding, does not produce better results for images that contain both foreground and background colours. To begin, features of the image under consideration are obtained based on luminance distribution. Binarization [34] was then performed using a decision-tree-based method that chose different features of color images to binarize images. If it is discovered that the colors in a color image are intense within a defined color range during the feature extraction process [35], an effective saturation is put in to the image. In addition, if the image colors in the foreground are more dominant, luminance is one of the most important parameters to consider. Finally, luminance was supposed to apply when the coloring of the image's background appears to be strenuous within a specified boundary, and saturation was supposed to apply when the number of pixels to limited luminosity would be less than 60; however, both luminance and saturation are enforced. The analyses reported in this article include 519 colour images in total. The majority of those are itemized receipts as well as name-card images. The suggested binarization method outperforms others in terms of shape as well as connected-component in this study.

As said by [17], detecting text present in image, frames, or videos is thought to be an important step in retrieving any multimedia information. The author proposed an improved algorithm for detecting, localizing, and retrieving side to side oriented textual data in image frames with degraded backgrounds. The proposed method is based on colour reduction techniques, an edge detection-based method, and text region localization [5] used a projection profile, which evaluates geometric attributes of color images. This same algorithm generates a series of textbox with a very simple background that are prepared to be fed into an OCR engine for consequent character recognition. Promising research findings for a set of image frames and videos illustrate the approach's effectiveness.

As stated in [25], a texture-based method was used to detect textual data in the image frame. Support vector machines are utilized to evaluate the textual characteristics of texts. Rather than using a different method to retrieve textual data features, the SVM [14] based classifier is given the strengths of the raw picture elements that comprise a textual outline. Another method involves using a uninterruptedly adaptive mean shift algorithm to analyse textual data and classify textual areas.

According to [21], provides information about textual information available in images and video frames for annotation, indexing, and image constructing. Identification, localization, tracking, fetching, enrichment, and identification of text data from an image are all processes involved in retrieving such information from an image. Variations in textual data due to various parameters, on the other hand, may cause problems in automatic text extraction.

Based on [10], a two-phase noise removing scheme based on a two-phase noise removing technique from images such as salt and pepper is presented. In the first phase, an adaptive median filter is used to identify picture elements that are most probably influenced by noise. In the second phase, the image is revamped again using a regularisation function that is applied to the selected noisy images. Edge perpetuation and noise reduction, as well as their regained image frames, provide a considerable improvement over nonlinear filter.

In [31] proposed that using local wiener filtering in conjunction with the wavelet domain is an operative image noise removal technique for non-severely degraded images. In this paper, the author proposes a doubly local wiener filtering algorithms-based method, which uses an elliptic directional window for various sub-bands to perform the calculation on the variances of signal for noisy wavelet coefficients, and the two other process of local wiener filtering are accomplished on the distortion containing images. The outcomes obtained after the experiment showed that the algorithm proposed in this work may improve the denoising performance.

As per [15], a novel adaptive method for binarization and improvement of tarnished images is presented. The approach described does not show any user-changeable requirement parameters and can easily handle image distortion, which is commonly caused by shadows, uneven illumination, low contrast, extremely signal-dependent noise, smear, and strain.

According to [36], a Morphological Component Analysis (MCA) procedure based on sparse expression of signals]. The morphological component analysis is built on the hypothesis that each signal's elemental behavior must be segregated, and that there is a dictionary for doing so using sparse representation. Following that, the pursuit algorithm for the sparse representation can be used to obtain the desired separation. The paper also includes several image content application results as well as some theoretical results that explain the separation process.

In [37] presents a novel method that is focused on the addition of the basis pursuit denoising (BPDN) algorithm along with total-variation regularization methodology used for the separating of texture features and cartoon parts from the image. The author suggests using two dictionaries for the representation of textures and natural scene parts respectively. Both dictionaries prompt sparse representations of image over single image content. The main use of the basis pursuit denoising gives a method for desired separation as well as noise removal. The separation

process is directed using a TV regularization scheme, also it is used for removing ringing artefacts. A highly improved numerical scheme describes a method to provide the solution for a combined optimization problem and several investigational outcomes that validated the fulfillment of the proposed algorithm.

As [42] proposed modeling of textured images using function and partial differential equation minimization. In this work, the image is decomposed into a summation of two procedures represented by u+v. where u represents a bounded variation procedure, while v is a procedure denoting the texture or noise. The algorithm proposed uses a differential equation and is simple to solve. It also explained the method can be used for texture discrimination of textures and texture segmentation.

In [18], gives a theory of Marr's primal sketch that integrates three components like texture model, a generative model along with image primitives and a Gestalt field. It also describes the meaning of "stretchability", which helps to divides image into texture and geometry and after the study and examining two different types of the model i.e., the detailed Markov random field model and the generative wavelet/sparse coding model s

3. Problem Statement and Motivation

As an active field of research, a large scientific community has concentrated on pattern recognition and computer vision, detection of textual data, and recognition of textual data. Text finding and apprehension may be considered a lively exploring domain, due to recent development of various portable devices and smart phone-based applications. Nowadays, text detection is a difficult and complicated task as there is a strong difference between textual and non-textual areas, as well as segregating every character from the frame of reference. This makes the textual data fetching process much more difficult to streamline. Furthermore, the intensity of illumination is the most important part of what makes textual data identification and recognition in ordinary scene images intricate.

The intensity of photos is often influenced by available darkness and different lighting in the setting, but intricate backgrounds of photos are usually obtained from outdoor images, making text fetching more difficult to automate. As a consequence, effective text detection filtering is necessary [20]. As a result, sub parts of images from the main image are retrieved and classified as textual as well as non-textual. The same would be done for sliding-windows of different sizes.

An efficient classifier that identifies the textual and non-textual parts from natural scene images [49] with less error classification is required to eliminate this repetitive process. The textual part that has been classified is then given to a deep neural network for character recognition. DNN requires optimal parameter (weight) selection, which is accomplished using an optimization algorithm.

4. Proposed Methodology

This section goes over the proposed methodology. As shown in figure. 1, the proposed strategy is divided into three main components: In the first segment, distortion is detached from the input image and then restored by utilizing a filtering mechanism. This will increase the visibility of the source images. To maintain the image variations within the images, the contrast of local image and gradient of local image methods have been deployed to the reconstructed images as mentioned throughout the second part. In the final section, image segmentation, extraction of features, and textual and non-textual data classification are carried out.

The proposed method is simple, robust, and provides significant results when compared with the existing technique. Table 1 details the schematic text retrieval methodology, which begins with a pre-processed image frame. source image is transformed throughout the pre-processing step, If this is a colourized version of the image, it would be transformed to grayscale. Now for contrast improvement, its pre-processed images get processed also with an adaptive contrast map.

The improved image would then be segregated using the marker-controlled watershed technique. The features of segmented images are then fetched by applying the Gabor's transform and the stroke width transform, and the extracted features are fed into a weighted nave bayes classifier to distinguish text and non-text images.

An emperor penguin optimal method is used to obtain error reduction together with a weighted nave classification algorithm. The letters inside the words were segregated as well as offered to a deep neural network with particle swarm optimization algorithm to final identification of characters in textual images.

The proposed method is simple, robust, and provides significant results when compared with the existing technique. Table 1 details the schematic text retrieval methodology, which begins with a pre-processed image frame. source image is transformed throughout the pre-processing step, If this is a colourized version of the image, it would be transformed to grayscale. Now for contrast improvement, its pre-processed images get processed also with an adaptive contrast map.

The improved image would then be segregated using the marker-controlled watershed technique. The features of segmented images are then fetched by applying the Gabor's transform and the stroke width transform, and the extracted features are fed into a weighted nave bayes classifier to distinguish text and non-text images.

An emperor penguin optimal method is used to obtain error reduction together with a weighted nave classification algorithm. The letters inside the words were segregated as well as offered to a deep neural network with particle swarm optimization algorithm to final identification of characters in textual images.
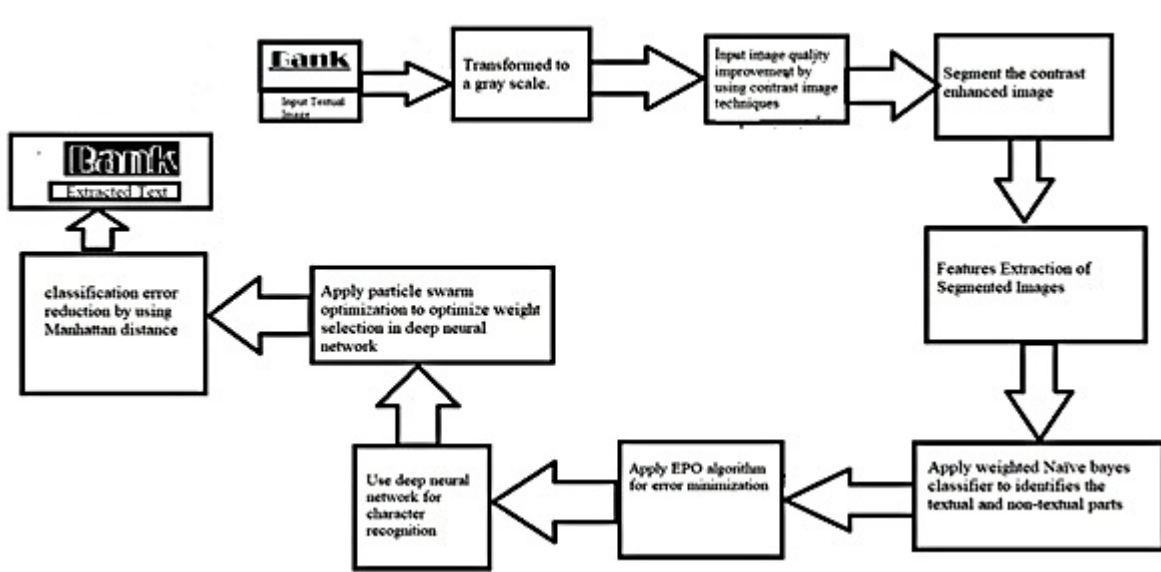
Figure 1. Work flow of suggested methodology for text extraction

Table 1. Algorithm of proposed methodology for entire   text extraction process

Input: Natural scene image with text Output: Extracted text
Step 1: Incorporate filtration technologies to enhance the smoothness, erase noise, and re-establish the image data.
Step 2: If the image to be input is in colour, this will be converted into grayscale.
Step 3: The input image is processed utilising gradient image and contrast image techniques.
Step 4: To enhance the intensity of a source images, an adaptive contrast map would be used.
Step 5: Segment the contrast enhanced image using marker-based watershed segmentation algorithm.
Step 6: Features present in segmented image is retrieved by using Gabor transform, and Stroke width transform.
Step 7: Based on this extracted feature, the weighted naïve bayes classifier identified the textual and non-textual parts.
Step 8: The error occur during the process of classification is minimized by using emperor penguin optimization algorithm by providing optimal solution and it also prevents the solution from falling into the local optimum.
Step 9: The classified textual part is then given for deep neural network for identification of character[9].
Step 10: Optimal parameter (weight) selection is necessary for deep neural network which is achieved by particle swarm optimization
Step 11: The classification error that occur during text extraction is minimise by determining the Manhattan distance between the strings.
Step 12: Accomplish a Lexicon search; if the Manhattan proximity is zero, the text or string is the same; otherwise, if the proximity is one, the optimised word is gained.

The suggested text retrieval process's work flow would be to properly recognise a text from an image, as well as the identified labels were saved in such a text document. The lexicon search was then used to compute the value of distance and if the computed distance measure is zero, the retrieved scene text and sequence are identical; if the distance is non-zero, the improved composition of word is identified. The process's main goal is to recognise a character through natural scenes and afterwards merge it to interpret a text and also make an appropriate classification model with few errors in order to increase a text classification process's reliability.

5. Results and Discussion

The suggested process is carried out, and also the results are evaluated to determine precision, recall, accuracy, and F1-score. After this, an assessed measurement has been contrasted to some of the most notable available methods. The above methodology has been discovered to be much more efficacious than other traditional approaches. Figure. 2 shows textual images used in this suggested technique. Figure. 3 illustrated images obtained after executing a pre-processing step and then watershed segmentation performed and segmented image is depicted in figure. 4. Correspondingly, a text extracted from such an obtainable source image is reported in figure. 5.

Figure 2. Textual images applied in suggested methodology

Figure 3. Pre-processed textual images after utilising gradient image and contrast image techniques.

Figure 5. Textual images after applying segmentation

Figure 4. Textual images after applying adaptive contrast map.

Figure 6. Extracted textual images after applying deep neural network and particle swarm optimization

The textual images used for this entire implementation stage are taken from the IIIT5K repository, which is widely used by various scientific professionals. All images throughout this repository vary greatly with scale, distortion, appearance and design with blur, color, distortion, font, and brightness.

The dataset used here is comprised of about 3,000 textual images taken from scenario photographs, including pictures with text-data from born-digital images. Of such 3000 pictures, 1000 will be chosen for model training, while 2000 will be picked for testing of the model. A few of the images captured from every phase for such an implementation phase as well as its corresponding outcome can be seen in the figure. 7.

Figure 7. Extracted textual images after applying DNN and PSO on four different input textual images

5.1 Evaluation metrics

Ansari et al [6], He et al [19], Almaz'an [4], Khlif [23], Zhu [53], Zhang [52], R-FCN, FasterR-CNN [32]. Are a few well-known textual data retrieval methodologies that are used to draw comparison [19]. The measurements of results are obtained with four True Positive (TP), True Negative (TN), False Positive (FP) and False Negative parameters. (FN).

The text part that are correctly identified as text is determined by TPThe text part that is incorrectly identified as non-text part is determined by FN.

The non-text part that is correctly identified as non-text is determined by TN. The non-text part that is incorrectly identified as text is determined by FP.

5.1.1 Accuracy

The most basic and important metric used in evaluate the effectiveness of classification and recognition is accuracy. The formula for determining the accuracy is given in equation (1).

$$Accuracy(in\%) = ((TP+TN)/(TP+TN+FP+FN)) \times 100 \quad (1)$$

Table 2. Computation of accuracy for proposed and prevailing methods

Methods	Accuracy (in percentages)
Proposed	98.39
Ansari et al [6]	95.3
He et al [19]	94
Almaz'an et al. [4]	93.66

As shown in table 2, the suggested technique has a better accuracy value compared to the existing method of text extraction. Few of the known methodology such as Ansari [6], He [19], and Almaz'an [4] are considered as a pre-existing method. Based on these values the below graph is plotted shown in figure 8. The accuracy of this work is found to be higher than the remaining three existing techniques.
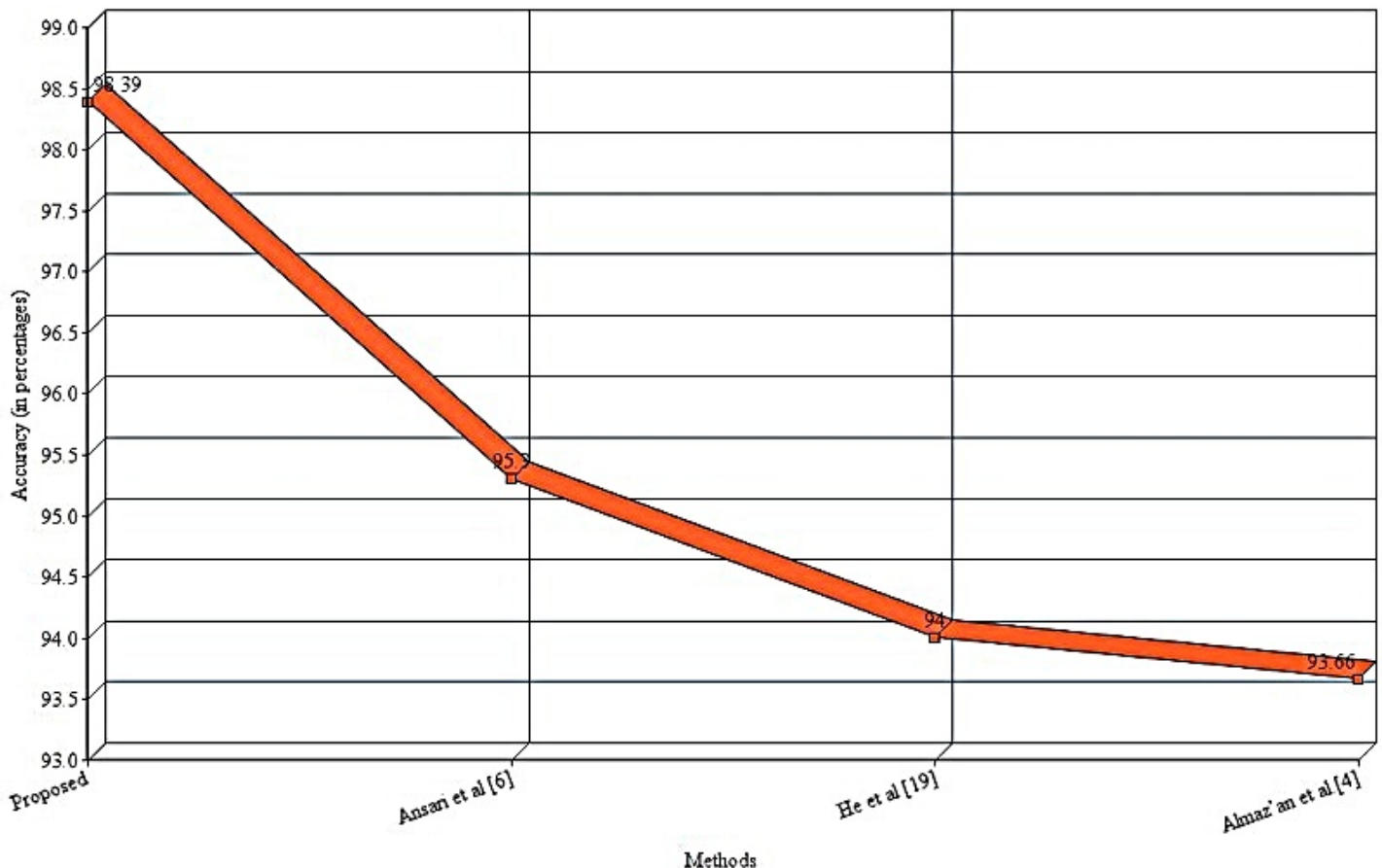

Figure 8. Accuracy (in percentages) of proposed and existing methods

5.1.2 Precision

Precision is calculated as the proportion of true positives (TP) to complete detections. Precision can be expressed mathematically in equation (2). As shown in table 3, the precision efficacy of the suggested approach is considerably higher than those of the existing method of text extraction. Few of the known approaches like like Khlif [23], Zhu [53], Zhang [52], R-FCN [12], FasterR-CNN FasterR-CNN [32] are considered as a pre-existing method. Based on these values the below graph is plotted shown in figure 9.

$$p = [TP/(TP+FP)] \times 100 \qquad (2)$$

Table 3. Precision of proposed and prevailing methods

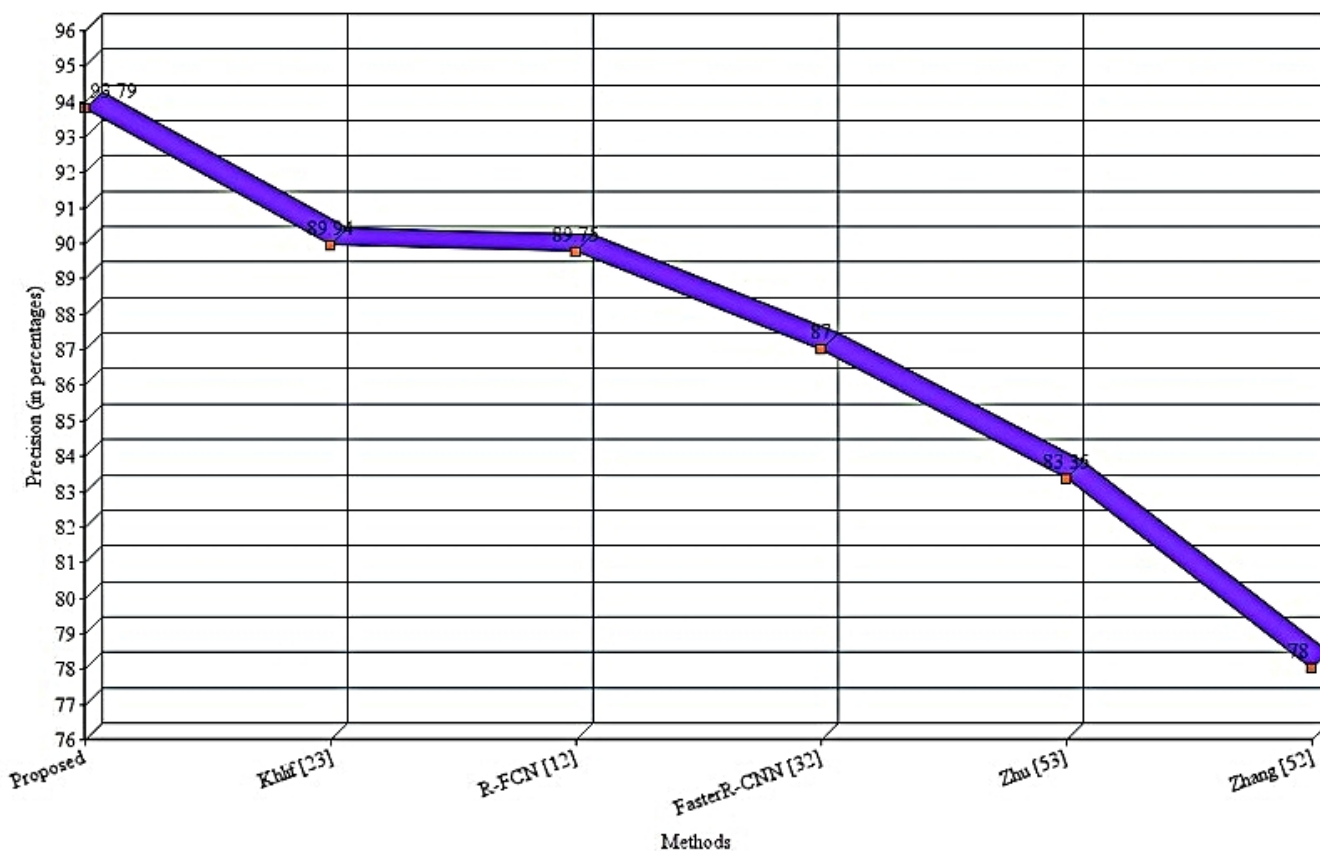| Methods	Precision (in percentages)
Proposed	93.79
Khlif [23]	89.94
R-FCN [12]	89.75
FasterR-CNN [32]	87
Zhu [53]	83.35
Zhang [52]	78

Figure 9. Comparison of precision for proposed and existing methods

5.1.3 Recall

The proportion of the reported true positive textual data to the entire identified true positive text as well as false negative textual data is determined by recall. The recall metric-score [37] can be mathematically expressed by equation (3).

.
$$r = (TP/(TP+FN)) \times 100 \qquad (3)$$

As shown in table 4, the suggested approach has an elevated recall value than that of the existing method of textual data fetching.
Few of the known approaches like like Khlif [23], Zhu [53], Zhang [52], R-FCN [12], FasterR-CNN FasterR-CNN [32] are

taken as the existing method. Based on these values the below graph is plotted shown in figure 10

Table 4. Recall of proposed and prevailing methods

Methods	Recall (in percentages)
Proposed	96.8
Khlif [23]	83.28
Zhu [53]	84
Zhang [52]	88
R-FCN [12]	76
FasterR-CNN [32]	75

Figure 10. Comparison of recall for proposed and existing methods

5.1.4 F1-score

Among a various performance parameter, the F1 score is considered an essential one. It functions as a metric for the proposed method. Precision and recall are both utilised in the F1-score estimation. The F1-score [42] can be expressed mathematically by equation (4). The F1-score value of suggested approach gives better outcome than the pre-existing method of text extraction as shown in the table 5.

Few of the known approaches like Khlif [23], Zhu [53] , Zhang [52], R-FCN [12], FasterR-CNN FasterR-CNN [32] are considered as a pre-existing method. Based on these values the below graph is plotted shown in figure 11.

F1-score (in percentage) = [2 * { (p * r)/(p + r) }] * 100   (4)

Table 5. F1-score for suggested and prevalent approaches

Methods	F1-score (%)
Proposed	95.2
Khlif [23]	85.94
Zhu [53]	84
Zhang [52]	83
R-FCN [12]	83
FasterR-CNN [32]	80

Figure 11. Comparison of F1-score for proposed and existing methods

5.2 Comparative Analysis

The precision, recall, and F1-score performance calculations of suggested and pre-existing textual data identification methods are depicted in figure 12. Textual data from rationally intricate corrupted images. The outcomes of the performance criteria being expressed as a percentage The

accessible alternatives show that the proposed methods outperform than other existing techniques for extracting
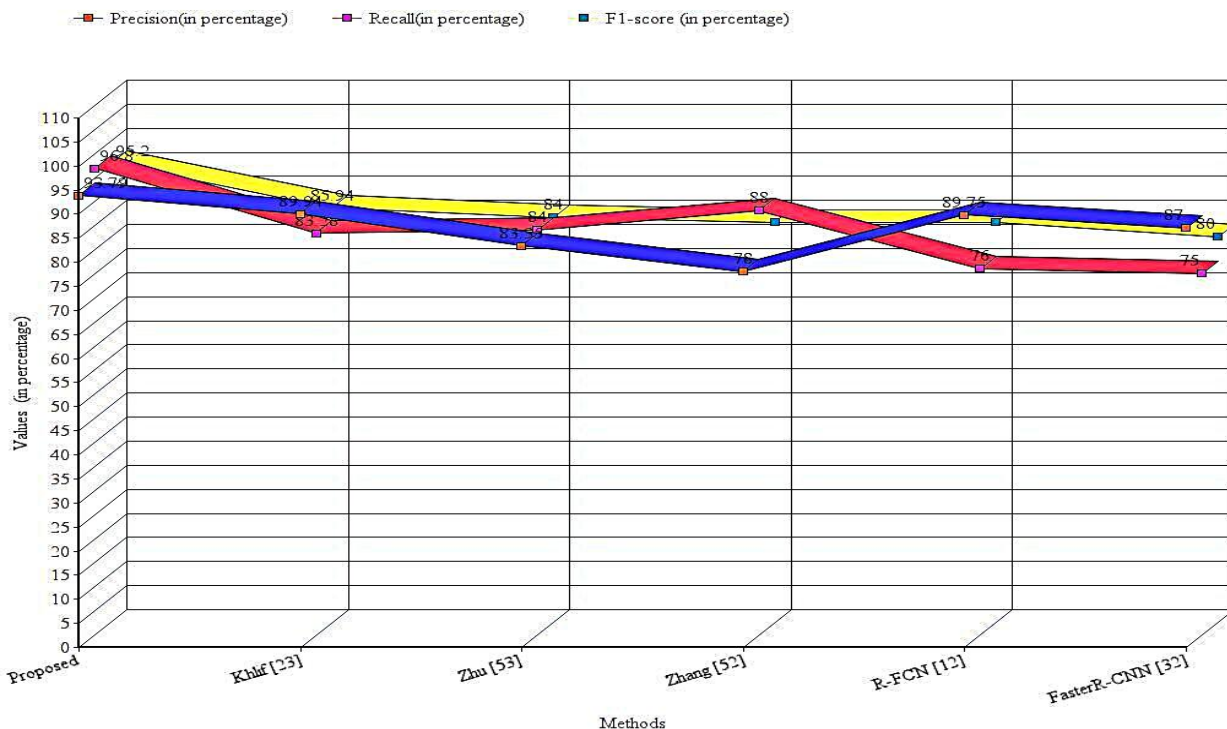

Figure 12. The comparative analysis Precision, Recall and F1-score

6. Conclusion

In this research, an effective method based on deep learning [38] is used to recover text from intricately corrupted images. To begin, smoothing, noise removal, and restoration techniques are applied to the input images, and if the input image is coloured, it is converted to a grayscale. The input image is then processed using gradient and contrast image techniques to regain the image's quality. Following that, the input image's contrast will be improved using an adaptive contrast map. Stroke width transform, Gabor transform, and weighted naïve bayes classifier techniques are also used to segment, extract features, and detect textual and non-textual components in complex degraded images.

Finally, a combination of deep neural network and particle swarm optimization is being used to recognise classified text. The dataset IIIT5K is used for the development portion, and while high performance is achieved with parameters such as accuracy, recall, precision, and F1 score, characters may occasionally deviate. Alternatively, the same character is frequently extracted [3] multiple times, which may result in incorrect textual data being extracted from natural images. As a result, an efficient technique for avoiding such flaws in the text retrieval process must be implemented in the near future. In order to improve the performance of the deep neural network architecture, an adaptive [46] rain optimization technique may be used.

References

[1]. Ahmed, S. B., Naz, S., Razzak, M. I., Rashid, S. F., Afzal, M. Z., & Breuel, T. M. (2016). Evaluation of cursive and non-cursive scripts using recurrent neural networks. Neural Computing and Applications, 27(3), 603–613. doi:10.1007/s00521-015-1881-4.

[2]. Ahmed, S. B., Naz, S., Razzak, M. I., & Yusof, R. B. (2019). A novel dataset for English-Arabic scene text recognition (EASTR)-42K and its evaluation using invariant feature extraction on detected extremal regions. IEEE Access, 7, 19801–19820. doi:10.1109/ACCESS.2019.2895876.

[3]. Ali, A., Pickering, M., & Shafi, K. (2018). Urdu natural scene character recognition using convolutional neural networks. In 2nd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), IEEE, 2018, March (pp. 29–34). IEEE Publications.

[4]. Almazán, J., Gordo, A., Fornés, A., & Valveny, E.. (2014). Word spotting and recognition with embedded attributes. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(12), 2552–2566. doi:10.1109/TPAMI.2014.2339814.

[5]. Andrew, T. D. (January 9, 1998). Representing multiple region of interest with wavelets. Proceedings of the SPIE, 3309, Visual Communications and Image Processing '98, 975.

[6]. Ansari, G. J., Shah, J. H., Yasmin, M., Sharif, M., & Fernandes, S. L. (2018). A novel machine learning approach for scene text extraction. Future Generation Computer Systems, 87, 328–340. doi:10.1016/j.future.2018.04.074.

[7]. Antonini, M., Barlaud, M., Mathieu, P., & Daubechies, I.. (1992). Image coding using wavelet transform. IEEE Transactions on Image Processing, 1(2, April), 205–220. doi:10.1109/83.136597.

[8]. Baran, R., Partila, P., & Wilk, R. (2018). Automated text detection and character recognition in natural scenes based on local image features and contour processing techniques. In. Advances in Intelligent Systems and Computing International Conference on Intelligent Human Systems Integration. Cham, Germany: Springer, 42–48. doi:10.1007/978-3-319-73888-8_8.

[9]. Bhunia, A. K., Kumar, G., Roy, P. P., Balasubramanian, R., & Pal, U. (2018). Text recognition in scene image and video frame using Color Channel selection. Multimedia Tools and Applications, 77(7), 8551–8578. doi:10.1007/s11042-017-4750-6.

[10]. Chan, R. H., Ho, C. W., & Nikolova, M. (2005). Salt-and-pepper noise removal by median-type noise detectors and detail-preserving regularization. IEEE Transactions on Image Processing, 14(10), 1479–1485. doi:10.1109/tip.2005.852196.

[11]. Chen, E. CT, & Chen, L. G. (1996, September). A self-adjusting weighted median filter for removing impulse noise in images. In Image processing. Proceedings., International Conference on (Vol. 1, pp. 419-422). IEEE Publications, 1996, [32] Peng, F., S., & Lucke, L. (1995, October). A hybrid filter for image enhancement.in icip (p. 163). IEEE Publications.

[12]. Dai, J., Li, Y., He, K., & Sun, J. (2016). R-fcn: Object detection via region-based fully convolutional networks. In Advances in Neural Information Processing Systems, 379–387.

[13]. Dutta, I. N., Chakraborty, N., Mollah, A. F., Basu, S., & Sarkar, R. (2019). Multi-lingual text localization from camera captured images based on foreground homogenity analysis. In Advances in Intelligent Systems and Computing, 149–158. doi:10.1007/978-981-13-1280-9_15

[14]. Francis, L. M., & Sreenath, N. (2019). Robust scene text recognition: Using manifold regularized Twin-Support Vector Machine. Journal of King Saud University – Computer and Information Sciences. doi:10.1016/j.jksuci.2019.01.013.

[15]. Gatos, B., Pratikakis, I., & Perantonis, S. J. (2006). Adaptive degraded document image binarization. Pattern Recognition, 39(3), 317–327. doi:10.1016/j.patcog.2005.09.010.

[16]. Ghai, D., & Jain, N. (2019). Comparative Analysis of Multi-scale Wavelet Decomposition and k-Means Clustering Based Text Extraction. Wireless Personal Communications, 109(1), 455–490. doi:10.1007/s11277-019-06574-w.

[17]. Gllavata, J., Ewerth, R., & Freisleben, B. (2003, September). A robust algorithm for text detection in images.in Image and Signal processing and Analysis, 2003. Proceedings of the 3rd International Symposium on, 2.IEEE p. 2003. Illinois School Psychologists Association.

[18]. Guo, C., Zhu, S., & Wu, Y. (2003). Towards a mathematical theory of primal sketch and Sketchability. In Proceedings of the

Ninth IEEE International Conference on Computer Vision (ICCV), (Nice, France), October 2003.

[19]. He, P., Huang, W., Qiao, Y., Loy, C. C., & Tang, X. (2016). Reading scene text in deep convolutional sequences. In Thirtieth AAAI conference on artificial intelligence.

[20]. Huang, Z., Zhong, Z., Sun, L., & Huo, Q. (2019). Mask R-CNN with pyramid attention network for scene text detection. In IEEE Winter Conference on Applications of Computer Vision (WACV) (pp. 764–772). IEEE Publications.

[21]. Jung, K., In Kim, K., & K. Jain, A. (2004). Text information extraction in images and video: A survey. Pattern Recognition, 37(5), 977–997. doi:10.1016/j.patcog.2003.10.012.

[22]. Khare, V., Shivakumara, P., Raveendran, P., & Blumenstein, M. (2016). A blind deconvolution model for scene text detection and recognition in video. Pattern Recognition, 54, 128–148. doi:10.1016/j.patcog.2016.01.008.

[23]. Khlif, W., Nayef, N., Burie, J. C., Ogier, J. M., & Alimi, A. (2018). Learning text Component Features via convolutional neural networks for scene text detection. In 13th IAPR International Workshop on Document Analysis Systems (DAS) (pp. 79–84). IEEE Publications.

[24]. Kumuda, T., & Basavaraj, L. (2017). Edge based segmentation approach to extract text from scene images. In 7th International Advance Computing Conference (IACC). IEEE Publications, Institute of Electrical and Electronics Engineers. (2017, January),pp. 706–710.

[25]. Kwang In Kim, Keechul Jung, & Jin Hyung Kim. (2003). Texture-based approach for text detection in images using support vector machines and continuously adaptive mean shift algorithm. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12), 1631–1639. doi:10.1109/TPAMI.2003.1251157.

[26]. Ma, J., Shao, W., Ye, H., Wang, L., Wang, H., Zheng, Y., & Xue, X. (2018). Arbitrary-oriented scene text detection via rotation proposals. IEEE Transactions on Multimedia, 20(11), 3111–3122. doi:10.1109/TMM.2018.2818020.

[27]. Manne, R., & Kantheti, S. C. (2021). Application of artificial intelligence in healthcare: Chances and challenges. Current Journal of Applied Science and Technology, 40(6), 78–89. doi:10.9734/cjast/2021/v40i631320.

[28]. Mehmood, Z., Mahmood, T., & Javid, M. A. (2018). Content-based image retrieval and semantic automatic image annotation based on the weighted average of triangular histograms using support vector machine. Applied Intelligence, 48(1), 166–181. doi:10.1007/s10489-017-0957-5.

[29]. Pandey, D., Pandey, B. K., & Wairya, S. (2021). Hybrid deep neural network with adaptive galactic swarm optimization for text extraction from scene images. Soft Computing, 25(2), 1563–1580. doi:10.1007/s00500-020-05245-4.

[30]. Paul, S., Saha, S., Basu, S., Saha, P. K., & Nasipuri, M. (2019). Text localization in camera captured images using fuzzy distance transform based adaptive stroke filter. Multimedia Tools and Applications, 78(13), 18017–18036. doi:10.1007/s11042-019-7178-3.

[31]. Peng-Lang Shui, P. L. (2005). Image denoising algorithm via doubly local Wiener filtering with directional windows in wavelet domain. IEEE Signal Processing Letters, 12(10), 681–684. doi:10.1109/LSP.2005.855555.

[32]. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems, 91–99.

[33]. Sain, A., Bhunia, A. K., Roy, P. P., & Pal, U. (2018). Multi-oriented text detection and verification in video frames and scene images. Neurocomputing, 275, 1531–1549. doi:10.1016/j.neucom.2017.09.089.

[34]. Sauvola, J., & Pietikäinen, M. (2000). Adaptive document image binarization. Pattern Recognition, 33(2), 225–236. doi:10.1016/S0031-3203(99)00055-2.

[35]. Sobottka, K., Kronenberg, H., Perroud, T., & Bunke, H. (2000). Text extraction from colored book and journal covers. International Journal on Document Analysis and Recognition, 2(4), 163–176.

[36]. Starck, J. L., Elad, M., & Donoho, D. (2004). Redundant multiscale transforms and their application for morphological component separation. Advances in Imaging and Electron Physics, 132, 287–348. doi:10.1016/S1076-5670(04)32006-9.

[37]. Starck, J. L., Elad, M., & Donoho, D. L. (2005). Image decomposition via the combination of sparse representations and a variational approach. IEEE Transactions on Image Processing, 14(10), 1570–1582, October 2005. doi:10.1109/tip.2005.852206.

[38]. Tang, Y., & Wu, X. (2018). Scene text detection using super pixel-based stroke feature transform and deep learning based region classification. IEEE Transactions on Multimedia, 20(9),2276–2288. Doi:10.1109/TMM.2018.2802644.

[39]. Tian, C., Xia, Y., Zhang, X., & Gao, X.. (2017). Natural scene text detection with MC–MR candidate extraction and coarse-to-fine filtering. Neurocomputing, 260, 112–122. doi:10.1016/j.neucom.2017.03.078.

[40]. Trémeau, A., Fernando, B., Karaoglu, S., & Muselet, D. (2011, April). Detecting text in natural scenes based on a reduction of photometric effects: Problem of text detection. In. Lecture Notes in Computer Science International Workshop on Computational Color Imaging. Berlin, Heidelberg: Springer, 230–244. doi:10.1007/978-3-642-20404-3_18.

[41]. Tsai, C. M., & Lee, H. J. (2002). Binarization of color document images via luminance and saturation colorfeatures. IEEE Transactions on Image Processing, 11(4), 434–451. doi:10.1109/TIP.2002.999677.

[42]. Vese, L. A., & Osher, S. J. (2003). Modeling textures with total variation minimization and oscillating pattern in image processing. Journal of Scientific Computing, 19(1/3), 553–572. doi:10.1023/A:1025384832106.

[43]. Wang, L., Uchida, S., Zhu, A., & Sun, J. (2018). Human reading knowledge inspired text line extraction. Cognitive Computation, 10(1), 84–93. doi:10.1007/s12559-017-9490-4.

[44]. Wang, Y., Shi, C., Xiao, B., Wang, C., & Qi, C. (2018). CRF based text detection for natural scene images using convolutional neural network and context information. Neurocomputing, 295, 46–58. doi:10.1016/j.neucom.2017.12.058.

[45]. Wang, Y., Wang, L., & Su, F. (2018). A robust approach for scene text detection and tracking in video. In. Lecture Notes in Computer Science Pacific Rim Conference on Multimedia. Cham, Germany: Springer, 303–314. doi:10.1007/978-3-030-00764-5_28.

[46]. Wu, H., Zou, B., Zhao, Y. Q., & Guo, J. (2017). 'Scene text detection using adaptive color reduction', adjacent character model and hybrid verification strategy. Visual Computer, 33(1), 113–126. doi:10.1007/s00371-015-1156-1.

[47]. Xue, M., Shivakumara, P., Zhang, C., Lu, T., & Pal, U. (2019). Curved text detection in blurred/non-blurred video/scene images. Multimedia Tools and Applications, 78(18), 25629–25653. doi:10.1007/s11042-019-7721-2.

[48]. Ye, Q., Huang, Q., Gao, W., & Zhao, D. (2005). Fast and robust text detection in images and video frames. Image and Vision Computing, 23(6), 565–576. doi:10.1016/j.imavis.2005.01.004.

[49]. Yin, X. C., Pei, W. Y., Zhang, J., & Hao, H. W. (2015). Multi-orientation scene text detection with adaptive clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1930–1937. doi:10.1109/TPAMI.2014.2388210.

[50]. Yin, X. C., Yin, X., Huang, K., & Hao, H. W. (2014). Robust text detection in natural scene images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 36(5), 970–983. doi:10.1109/TPAMI.2013.182.

[51]. Yuan, Q., & Tan, C. L. (2001). Text extraction from gray scale document images using edge information. In Document Analysis and recognition. Sixth International Conference on, 2001.Proceedings (pp. 302–306). IEEE Publications.

[52]. Zhang, Z., Zhang, C., Shen, W., Yao, C., Liu, W., & Bai, X. (2016). Multi-oriented text detection with fully convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4159–4167).

[53]. Zhu, A., & Uchida, S. (2017). Scene text relocation with guidance. In 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 1 (pp. 1289–1294). IEEE Publications.

[54]. Abdulsahib, G. M., & Khalaf, O. I. (2018). Comparison and Evaluation of Cloud Processing Models in Cloud-Based Networks. International Journal of Simulation-- Systems, Science & Technology, 19(5). 64.

[55]. Osamah Ibrahim Khalaf, Ghaida Muttashar Abdulsahib and Muayed Sadik, 2018. A Modified Algorithm for Improving Lifetime WSN. Journal of Engineering and Applied Sciences, 13: 9277-928