Design of Fractional Poisson Enhancement Model for Text Detection and Recognition in Video Frames

[1]Laxmikant Eshwarappa , [2]Dr. G. G. Rajput

Assistant Professor, Dept. of MCA
Sharnbasva University, Kalaburagi, Karnataka, India
Professor, Dept. of Computer Science
Akkamahadevi Women's University, Vijayapur, Karnataka, India
elaxmikant@gmail.com ,ggrajput@yahoo.co.in

ABSTRACT:

In order to accurately identify and recognize text, it is common practice to employ the Laplacian operation on video images to boost contrast. While Laplacian operation does enhance contrast, it also adds an excessive amount of noise, which is a fact. To alleviate this, the existing approaches provide different filters and enhancement methods. In this paper, we propose a generalized enhancement model based on fractional calculus to increase the quality of images generated by Laplacian procedures. By considering edges and their neighbor information, the recommended method creates a mathematical model for enhancing low-contrast information in video and scene photographs. Experimental results on several datasets show that the proposed enhancement model significantly enhances the accuracy of text recognition and identification approaches. When the enhancement model is contrasted with traditional enhancement models, the proposed method outperforms the present models in terms of quality metrics. The effectiveness of the proposed model is validated using text detection and recognition tests.

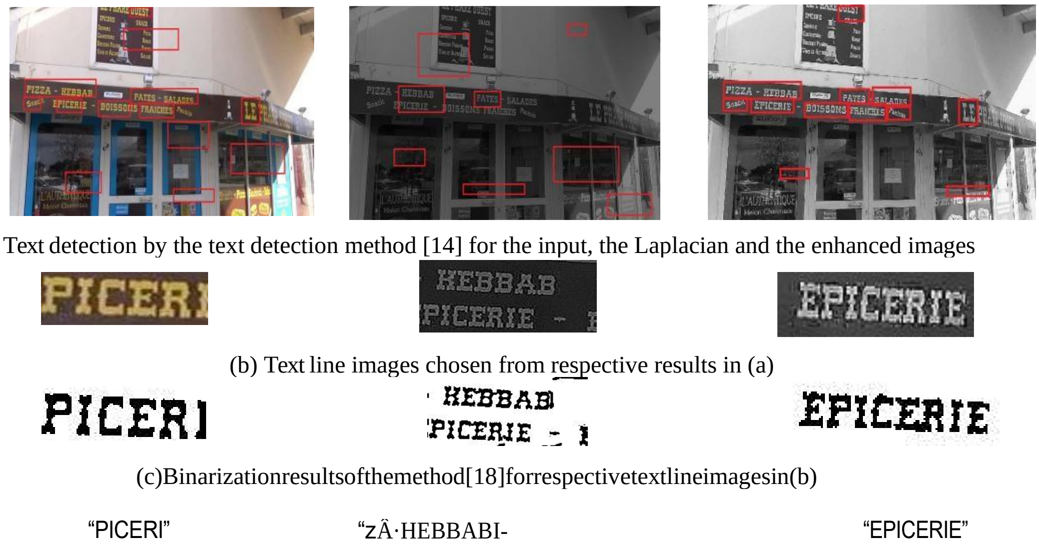
Keywords: Text Detection and Recognition in Video Frames, Laplacian operation, Average PSNR and SSIM for enhanced images, Fractional Poisson Enhancement

DOI: 10.48047/ecb/2023.12.Si8.630

Introduction and Background

As systems for assisting a blind person to walk freely on roads, safely driving, and tracking license plates of moving vehicles, real-time applications like these frequently require more than 90% detection and recognition accuracies [1-5], researchers are paying more attention to video text detection and recognition with the aim of improving the performances of the existing text detection and recognition methods. As a result of motion blur, lighting, uneven illumination, text motions, and complicated backgrounds, video pictures dramatically degrade, making it difficult for researchers to achieve such high accuracy [3, 6, 7].To solve these issues, methods based on gradient operation with Laplacian mask have been proposed in the literature [8–13] for improving text information in video images. Laplacian aids in identifying abrupt changes from background to foreground and vice versa by providing high positive and negative peaks. Since this data serves as the foundation for extracting features to identify text and distinguish foreground (text) from background in binarization, it is helpful for both text detection and identification. To separate the words in each text line in a film, Shivakumara et al. [10] employed strong positive and negative peaks. In order to recognize text in video pictures, Phan et al. [11] exploited the transition from background to foreground and vice versa.

Similar to this, text in video or photos has been binarized and recognized using the Laplacian operation [12, 13]. It is true that Laplacian aids in improving text information and text separation; nevertheless, when applied to a picture, Laplacian creates excessive noise. The previous solutions [14–17] often provide various criteria based on filters to eliminate the noise impact brought on by Laplacian operation in order to solve this issue. It is clear from the subsequent techniques. In order to reduce the impact of noise on the selection of text candidates, Shivakumara et al. [14] suggested wavelet and color features for text identification in video, using an augmented picture created by combining the R, G, and B color spaces as the input.Shivakumara et al[16] .'s Laplacian strategy for multi-oriented text identification in video also included a proposal for the Fourier transform as the appropriate low pass filter to eliminate disturbances brought on by Laplacian operation. Fig. 1 tests a text identification approach that already exists [14] that makes use of wavelet and color characteristics, and applies automated parameter tweaking for binarization [18] before and after Laplacian with the aid ofPublicly accessible Optical Character Recognizer (OCR) [19]. As seen in Fig. 1, the text identification algorithm for the photos in (a) recognizes texts, but produces more false positives for the input image than for the improved image.However, due to noise impact, the same text identification system misses certain characters and produces more false positives for the Laplacian picture than the input image. In contrast to both the input and the Laplacian pictures, the text identification technique correctly finds messages and produces fewer false positives for the improved image provided by the proposed model (it will be explained later in the suggested methodology section).The similar conclusion can be reached from the findings displayed in Fig. 1(b)-(d), where the OCR engine successfully detects the texts in the improved picture but misses a few letters for the texts in the input image and provides trash values for the texts in the Laplacian image. This demonstrates that a broader model is required to eliminate such operational effects.



(a) Text detection by the text detection method [14] for the input, the Laplacian and the enhanced images

(b) Text line images chosen from respective results in (a)

(c)Binarizationresultsofthemethod[18]forrespectivetextlineimagesin(b)

(d) Recognition results of the OCR engine [19] for the respective results in (c)

Figure1:(a)Text detection by detection method..(b)Text line images choosen from respective values of (a) …(c)Binary results…(d) Recognition results

We may conclude from the description above that there is no reliable improvement technique for minimizing the noise effect of Laplacian operation. As far as we are aware, there is currently no generic enhancement model for the distortions brought on by gradient or Laplacian operations.

1. Related Work

Over the past ten years, a number of techniques for text detection and recognition in video have been developed [1, 2]. The present techniques may be roughly divided into three categories: (1) connected component-based, (2) texture-based, and (3) gradient and edge-based techniques. The features of character components are explored for text detection using connected component based approaches, which are straightforward. Character shape analysis is always necessary since these approaches rely on the traits of character components. However, it is challenging to obtain precise forms for character components because of the Laplacian operation's distortion impact, poor resolution, and complicated video backdrop. This means that the effectiveness of these approaches for video text detection may be questionable. For instance, Rong et al. [20] suggested a two level approach based on the features of character components for word identification in natural scene photos. With the use of stroke width distance-based features and maximally stable extremal regions (MSER), Chen et al. [21] suggested a technique for reliable word identification in natural scene photos. Robust text identification in natural scene photos based on MSER, clustering, and character classifier was proposed by Yin et al. [22]. In order to categorize character components as text candidates, the approach analyzes their properties for the MSER output. To find real text candidates, single link clustering and character classifier are utilized.Since the methods mentioned above presume that a picture has strong contrast, they produce character components. The performance of the approach suffers greatly when used to video images because to the poor contrast and resolution. These techniques are therefore vulnerable to poor contrast, distortions, and complicated backgrounds.

Many approaches have been put forth employing texture characteristics for text identification in video to solve the issue of complicated backgrounds [1, 2]. These techniques take into account the text's visual pattern as a unique textural attribute. The approaches, however, provide poor accuracy for texts with a variety of typefaces and font sizes since establishing the texture attribute for text components is difficult.Additionally, to increase the accuracy of text detection, the majority of approaches utilize costly classifiers and extract a high number of features. They are therefore expensive to compute for real-time applications. For instance, Wavelet, Color Features, Fourier with Color Spaces, and Fourier with Lapalacian were suggested by Shivakumara et al. [14, 16, 23] for text identification in video.

These three techniques are effective for low contrast photos, but they need expensive transformations, which adds to their computing cost. Additionally, when an input picture incorporates distortion brought on by operations and movements, the approaches' performances suffer.

The use of gradient and edge information in text identification in video has been suggested as a way to reduce the computing burden. These techniques are effective for text identification in video because edge and gradient information gives important clues like strong gradient values for text pixels and crucial details of the character components residing in vertical and horizontal edge directions. Due to their simplicity and efficiency in gradient and edge operations, they are therefore more often used than linked component and texture-based approaches [1, 2]. Therefore, to enhance their performances, the majority of cutting-edge algorithms analyze gradient and edge information in various ways. To recognize writing in photographs of natural scenes, Epshtein et al. [24] suggested using canny edge images and the stroke width transform (SWT).In order to recognize text in videos, Phan et al. [11] devised a Laplacian technique that combines k-means clustering with Laplacian analysis. Multi-oriented video scene text identification using a Bayesian classifier and boundary growing was suggested by Shivakumara et al. [15]. However, this approach receives low accuracy since it frequently returns false positives for complicated backdrop photos. A novel approach for extracting text from videos based on stroke information was suggested by Huang et al. [17]. The technique obtains the edge map of a gradient picture made up of text rows. The technique then retrieves a character stroke image of text rows using the Laplacian image and the edge map.The method's primary flaw is its reliance on the geometrical characteristics of character components, as it is sometimes challenging to produce character components free of disconnections in situations with complicated backgrounds and low contrast. Additionally, the method's effectiveness suffers when dealing with distorted pictures.

Gradient and edge-based approaches are believed to be susceptible to distortions brought on by gradient operations as well as noises produced by other sources such non-uniform lighting and motion blur, while being computationally affordable and producing decent results. Additionally, several techniques employ various standards for eliminating the sounds produced by gradient operations and distortions. The approaches become inconsistent as a result of this. A generalized enhancement model might therefore be built to lessen the impact of distortions caused by gradient operations, particularly the Laplacian operation. Similar to this, other approaches for text recognition have been presented in literature [1]. The techniques now in use, as reported in the literature, address this issue in two ways: (1) Text recognition using custom features and classifiers that, in most cases, do not rely on binarization or readily accessible OCRs. In other words, they create a unique OCR for text recognition in videos. (2) Recognizing text in video by creating a reliable binarization technique that enables use of the existing OCR.The former is too expensive and comes with its own restrictions, such as the need for a classifier and the need for sample training, which limit the capacity to adapt from various scripts, data, and applications. The latter, however, is less costly than the former since it makes use of accessible OCRs. Therefore, rather than creating a separate OCR, which is not recommended, the present study concentrates on the latter to tackle the video text recognition problem.

As an illustration, Roy et al[25] .'s approach of binarization-based text recognition relies on the idea of fusion. Although this approach is effective for video texts, it cannot handle arbitrary orientations. A multi-scale framework for adaptive binarization of damaged document pictures utilizing Otsu thresholding was put out [26] by Moghaddam and Cheriet. Otsu thresholding, however, only works well when the intensity values are distinct from the backdrop. It's possible that video defies this restriction. By using the proper binarization techniques, Chattopadhyay et al. [27] suggested a strong OCR for identifying texts in various document pictures.But selecting one of these binarization techniques requires careful consideration of the image's content. A document binarization technique with automated parameter tweaking was presented by Howe [18]. This approach suggests an automated process to choose parameter values based on training and sample data. Text localization, augmentation, and binarization in multimedia documents were suggested by Wolf et al. [28]. Laplacian energy was another idea put up by Howe [13] for document binarization. The technique investigates binarization's Laplacian features, such as the change from background to foreground and vice versa. Document binarization using topological clustering-guided Laplacian energy segmentation was suggested by Ayyalasomayajula and Brun [12].

In conclusion, the aforementioned talks show that the majority of binarization techniques utilise transition data from background to foreground and vice versa. to distinguish between text and non-text pixels. At the same time, we can see that these strategies provide various filtering ways to lessen the noise impact caused by Laplacian operation. This demonstrates that there is no universal model that can be used to lessen the impact of Laplacian operation. As a result, the main goal of this study is to provide a generalized enhancement model for minimizing the noise impact.

3.Proposed Model

As can be seen from the literature study covered in the preceding section, Laplacian features are used by the majority of text detection and recognition systems to enhance their performance. We thus use the Laplacian sounds as a case study in this work to design an enhancement model. The key benefit of this model is that it was created to improve Laplacian operation output rather than taking into account manually inserted sounds as was done in previous research [29, 30]. We may thus infer that the suggested model can be applied as an augmentation method for distortions brought on by other sources.

The suggested approach is divided into two components. The suggested fractional Poisson enhancement model is summarized in Section 3.1, and the derivations and examples for the fractional Poisson model are provided in Section 3.2.

3.1 Overview of the Fractional Poission Model

The physical and engineering disciplines frequently employ fractional calculus and its applications. Additionally, fractional differentiation is regarded as a superb method for describing the fundamental characteristics of a variety of materials and processes [29, 30]. Numerous fractional calculus operators, including the Grünwald-Letnikov, Erdélyi-Kober, Caputo, Weyl-Riesz, and Riemann-Liouville operators, have been produced over the past 50 years [29, 30]. Fractional calculus has drawn a lot of interest in the realm of image processing for picture texture improvement and image denoising. All of the results based on fractional calculus operators demonstrated the effectiveness and dependability of these models, leading to high levels of permanent immunity against various noise forms [29, 30].The reasoning for picture enhancement based on fractional Poisson is that the images produced by the Laplacian operation will be introduced simultaneously with small changes in the grayscale. The nonlinearity of fractional Poisson, however, preserves high-frequency marginal characteristics in regions where changes in gray level are large, and enhances low-frequency details in regions where changes in gray level are negligible. We use fractional calculus to improve the quality of the pictures produced by the Laplacian operation as a result of this insight.

Images are processed using image enhancement techniques so that the finished product is more suited for a certain purpose than the original picture. In order to improve the quality of the pictures produced by the Laplacian operation, we offer a generalized enhancement model based on fractional Poisson in this study. Here, picture enhancement refers to altering the original values of digital pixels to produce targets and backgrounds that contrast more effectively.

a. Construction of Fractional Mean

The Riemann-Liouville fractional differential operator of the order used in our inquiry $0 < \alpha < 1$ [31]

$$D^{\alpha} f(t) = \frac{d}{dt^{a}} \int^{t} \frac{(t-\tau)^{-\alpha}}{\Gamma(1-\alpha)} f(\tau) d\tau. \quad (1)$$

The fractional integral operator for a continuous function f is represented by the equation below (t)
of the order $\alpha > 0$:

$$I^{\alpha}_{a}(t) = \int_{a}^{t} \frac{(t-\tau)^{\alpha-1}}{\Gamma(\alpha)} f(\tau) d\tau. \quad (2)$$

A fractional non-Markov Poisson stochastic process has been created using the Riemann-Liouville fractional differential operator based on the fractional extension of the Kolmogorov-Feller equation [32]. The probability distribution function P(x,sfractional)'s Kolmogorov-Feller equation is defined by

$$\frac{\partial P(x,s)}{\partial s} = \int_{-\infty}^{\infty} dy\omega(y)[P(x-y,s) - P(x,s)],$$

$$(x,0) = \delta(x),$$

where is the length y probability density. Additionally, the step length is randomly distributed according to where s is the order's time steps.

$$\Psi(s) = \frac{\sin \pi\alpha}{\pi}\int_{0}^{\infty} \frac{e^{-\rho s}d\rho}{2\cos(\pi\alpha)+\rho^{\alpha}+\rho^{-\alpha}}, \quad 0 < \alpha \le 1,$$

for which a fractional Poissonian distribution is used.

P(n,r) denotes the likelihood that there are n items in position r. The normalizing condition is met by probability P.

when condition

$\sum^{\infty} P(n,r) = 1.$

The following specific variant of the fractional Kolmogorov-Feller equation often expresses the probability P(x,s) [32]

$$P_{\alpha}(n,r) = \frac{(\bar{r}^{\alpha})^{n}}{n!}\sum_{k=0}^{\infty}\frac{(k+n)!}{k!}\frac{(-\bar{r}^{\alpha})^{k}}{\Gamma(\alpha(k+n)+1)}, \ 0 < \alpha \le 1. \quad (3)$$

As a result, the mean n of the fractional Poisson process may be easily determined as shown below.:

$$\bar{\bar{n}}_{\alpha} = \sum_{n=0}^{\infty} n P_{\alpha}(n,r) = \frac{\bar{r}^{\alpha}}{(\alpha+1)}, \quad (4)$$

where is the picture's mean value, which reflects the average of all its pixels, is the fractional power, and r is a tuning parameter used to raise the amount of visual contrast. In order to extend the dynamic range of the gray level, we change the mean in formula (4) to be the fractional mean m of the Laplacian picture (the images produced by the Laplacian operation). It is the Laplacian picture, while Ie is the created enhanced image. In this case, since is a real number, I is just multiplied by a real number. The fractional mean of the Laplacian picture is used to make the contrast enhancement of the Laplacian image reliant on it. In this, m is made adaptive, meaning that it has a lower value for dark pixels while having a larger value for light pixels.Both the value of and the content of the picture affect the fractional mean. Here, the concept of adaptively boosting gray level dynamic range is introduced.

Generally speaking, image contrast refers to the distinction between an object's visual attributes and its backdrop in an image.

Following are the steps for the suggested fractional picture enhancement algorithm:

Both of the parameters (, r) are quite helpful for raising an image's contrast to a higher degree. From (5), it is clear that the power of has increased both the fractional mean and the tuning parameter. As a result, various and r values will have an impact on various intensity areas in a particular Laplacian picture. The PSNR behavior of the resulting enhanced picture for various values of and r is shown in Fig. 2. Lower values of will result in a smaller PSNR by extending the dynamic range of dark pixels.Higher values of, on the other hand, will widen the dynamic range of bright pixels, which results in a sharp drop in PSNR. In its most basic version, we manually select the ideal values of and r. We select = 0.6 with the tuning parameter r=0.04 from Fig.2. (Dash-dot line blue color).

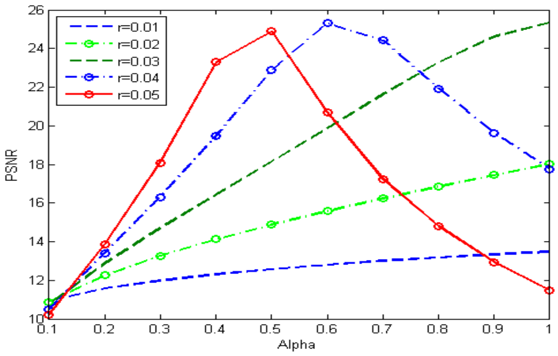
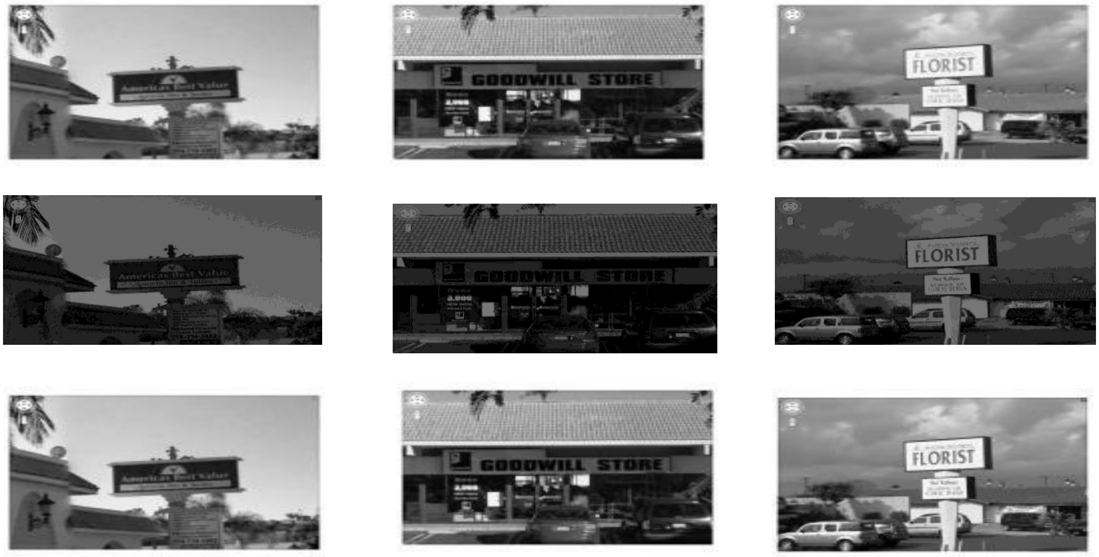


Figure 2: Optimal values of α and r manually

Figure 3 illustrates how the proposed model works by showing various input photos, corrupted Laplacian images from the Laplacian operation, and improved images produced by the proposed model. The improved pictures seem brighter than the input and Laplacian images, as can be seen in Fig. 3. Additionally, unlike the Laplacian picture, the improved image is noise-free. This improves the accuracy of text detection and identification techniques. The benefit of the suggested model is this.



(c) Enhanced images by the proposed model

Figure 3: **where (a) shows different input images, (b) denotes Laplacian images corrupted by the Laplacian operation,and (c) gives enhanced images given by the proposed model**

4.Experimental Results

The ICDAR 2013 scene dataset [33] and the Street View Data (SVT) [34] both contain high resolution, complex backgrounds and mostly horizontal texts, respectively. The ICDAR 2013 video dataset [33] and the ICDAR 2013 scene dataset [33] both contain texts with low resolution, complex backgrounds, different fonts or font sizes, and different orientations, respectively. In this study, we experiment with 6366 text graphics in total.The fundamental benefit of using these datasets is that they all offer the necessary ground truth for computing measures automatically. We employ industry-accepted metrics, such as Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM), which are often used to assess the quality of pictures [29, 30], to evaluate the improved images provided by the proposed model.In order to demonstrate the suggested model's performance in terms of quality indicators, we furtherContrast-Limited Adaptive Histogram Equalization (CLAHE), which is used to improve the contrast of grayscale images, Adjust Intensity Values to Specified Range (AIV), which is used to increase the contrast of an output image by mapping the intensity values of the input grayscale image, are a few examples of benchmark enhancement techniques that can be used to compare the results of the proposed model.The selection of these three procedures was motivated by the fact that they served as the foundation for the majority of improvement methods suggested in the literature [29, 30], either directly or indirectly. In order to demonstrate that the suggested model is correct and effective, we compare it with these basis.

We test the proposed enhancement model's performance on three standard noisy images, Lena, Cameraman, and Boat, by adding Gaussian noises with a standard deviation of 15 [29] because the noises produced by Laplacian operation resemble denoising. The results and discussions of this test are provided in Section 4.1.

We propose to undertake tests on text identification and recognition on the aforementioned data to demonstrate the utility of the suggested enhancement model as our datasets lack ground truth for noisy pictures (Laplacian images).In comparison to the input and Laplacian photos, we anticipate that text detection and identification results for the upgraded images should perform better.As a result, we implement the existing text detection methods, such as the Epshtein et al. method [24], which uses the stroke width transform to detect text in images of natural scenes, the Shivakumara et al. method [14], which suggests using wavelet and color features to detect text in videos, the Shivakumara et al. method [23], which suggests using the Fourier transform and RGB color space to detect text in videos, and the Shivakumara et.According to Chen et altechnique .'s [21], maximum stable extremal regions (MSER) are suggested for text candidate identification, and geometrical features are subsequently suggested for the removal of erroneous text candidates in photos of natural scenes. By suggesting adjustments to MSER and single link clustering to create text candidates, Yin et altechnique .'s [22] expands Chen et almethod .'s [21] or enhancing text identification outcomes for natural scene photos.

For determining weights and text candidate categorization, a character classifier and distance metric learning have been proposed. Similar to this, we implement binarization methods for recognition experiments, such as the Roy et al. method [25] that uses wavelet and gradient fusion for binarizing video text lines, the Chattopadhyay et al. method [27] that proposes automatic selection of binarization methods for different portions of a document, the Wolf et al. method [28] that uses edge and gradient information for binarization, and the Moghaddm and Cheriet's.Since recall, precision, and f-measure are the industry standard metrics for text detection, we utilize them to assess how well the text detection systems work. The binarization techniques are assessed using the recognition rate at the character level provided by the OCR engine in [19].

A. Evaluation of the Proposed EnhancementModel

Equations (7) and equation (8) describe the quality measurements that we employ to gauge the effectiveness of the suggested improvement strategy (9). The mean squared error between the corresponding pixel values of the Laplacian image (C) and the original picture is used to compute PSNR.

(O) [29, 30]:

$$\mathrm{PSNR}=10\log\frac{\max(C,O)^{2}}{\mathrm{MSE}} \qquad (6)$$

$$\mathrm{MSE} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\left(C(i,j)-O(i,j)\right)^{2} \qquad (7)$$

where max is the highest pixel value that the picture may contain. This value equals 255 when shown in a grayscale picture. The relationship between PSNR and MSE is inverse, measured in decibels. A higher number for PSNR suggests that the pictures are more comparable.

Similar to how SSIM is defined in equation (9). In order to compare the structural similarities and differences between the original and Laplacian pictures, the SSIM is introduced. According to its definition:

$$SSIM(x,y)=[\ l(x,y)]^{\alpha}.\ [c(x,y)]^{\beta}.[s(x,y)]^{\gamma} \qquad (8)$$

The structure comparison function is called s, the contrast comparison function is called c, and the luminance comparison function is called l. The relative significance of the three components is altered using the parameters,, and. According to the instructions in [36], where the values of the three parameters are regarded as larger than zero, we compute the values.

We calculate the two quality measures for several combinations, such as the enhanced image with the Laplacian image, the original image with the Laplacian image, and the enhanced image with the original image, as there is no ground truth for the input Laplacian pictures. PSNR and SSIM should provide high scores for an enhanced image with its original image, but lower scores for an enhanced image with its Laplacian image as compared to the original image with the Laplacian image. The enhanced picture with the original image (input image) scored better on both measures, but the enhanced picture with the Laplacian image scored lower, according to Table 1's quantitative results for the recommended approach.This finding may be evaluated combining the scores of the original picture (input image) with the Laplacian image since the scores for this combination are neither higher than those of Enhanced-Original (input) nor lower than those of Enhanced-Laplacian. We may conclude that the recommended improvement model shows promise as a consequence.

Fig.3 displays some representative qualitative findings of the proposed and current approaches, including (a) HE results, (b) CLAHE results, (c) AIV results, and (d) results of the proposed model. The data in Fig. 5.3 make it difficult to clearly discern any differences. However, we can see the differences when we compare the quantitative outcomes of the proposed and existing methodologies, which are provided in Table 1. Table 1 shows that, for all the datasets taken into consideration for testing, the PSNR results of the proposed algorithm beat those of the current approaches, namely, HE, CLAHE, and AIV.This is because the suggested technique has the capacity to automatically adopt appropriate picture enhancement settings by using the fractional mean of the input Laplacian image as an intensity transform to expand the dynamic range of the Laplacian image's gray levels. The comparison demonstrates that the suggested approach may be used to improve contrast in images with greater success than alternative methods.

Table 1: Average PSNR and SSIM for enhanced images with noisy and input images

Dataset	Number of images	PSNR			SSIM		
		Enhanced -Laplacian	Original-Laplacian	**Enhanced –Original (proposed)**	Enhanced –Laplacian	Original-Laplacian	**Enhanced –Original (proposed)**
ICDAR 2013 Video	5487	11.95	13.15	**24.26**	0.74	0.76	**0.95**
ICDAR 2013 Scene	229	12.10	13.39	**22.82**	0.74	0.76	**0.91**
SVT	350	12.90	15.15	**24.56**	0.74	0.79	**0.96**
MSRA	300	11.74	12.7	**26.99**	0.74	0.76	**0.98**
Average	Total = 6366	12.0	13.1	**24.2**	0.74	0.76	**0.95**

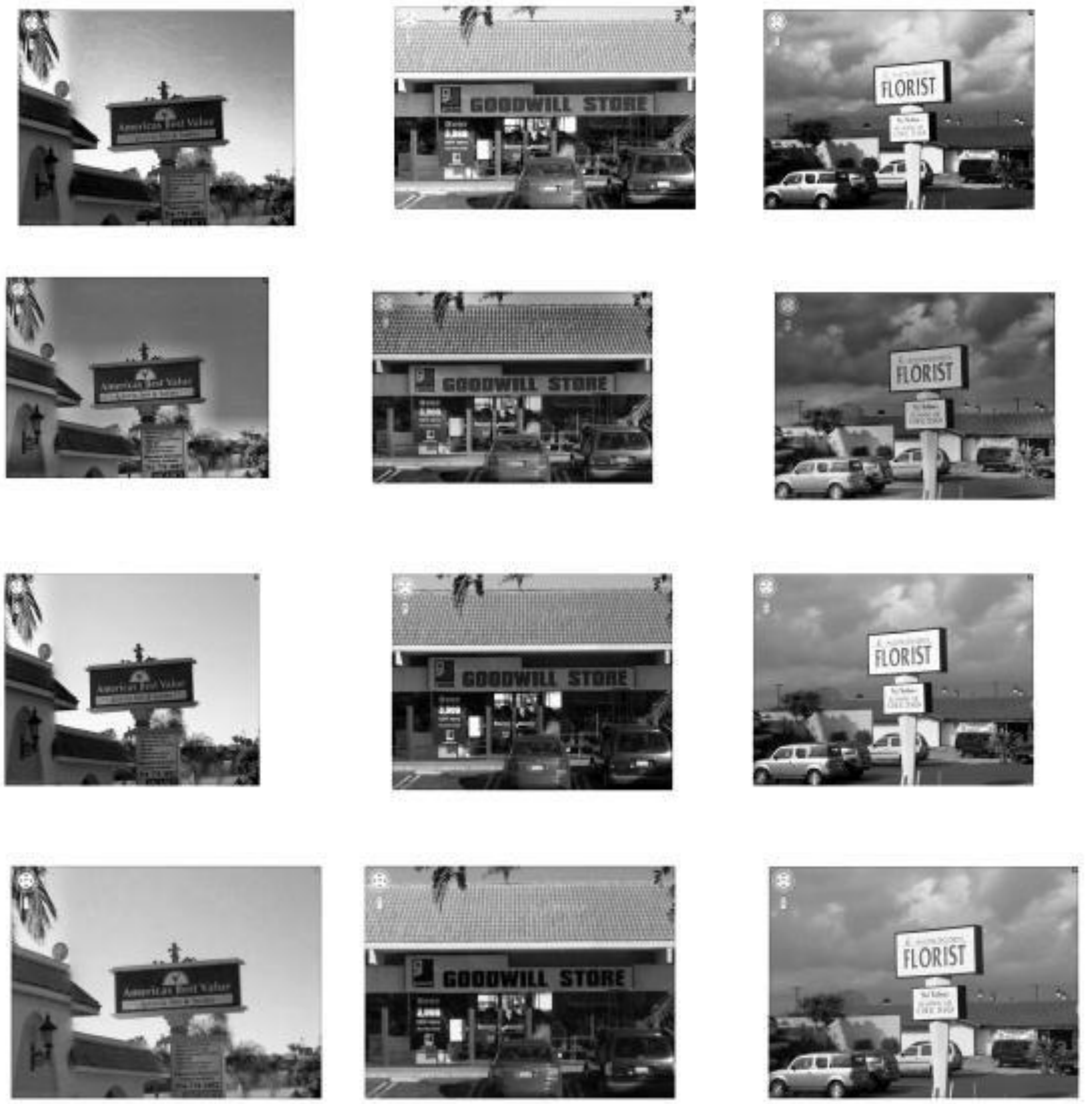


Figure 3:(b) Enhanced Results of the

Table 2: Quality measures of the proposed and existing models (we calculate PSNR and SSIM for Enhanced images with Original (input) images)

Dataset	HE		CLAHE		AIV		Proposed Model	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ICDAR 2013 Video	18.1	0.73	18.1	0.78	17.7	0.78	24.3	0.95
ICDAR 2013 Scene	14.8	0.63	14.8	0.63	14.8	0.63	22.8	0.91
SVT	15.8	0.72	15.8	0.72	15.8	0.72	24.5	0.96
MSRA	16.6	0.75	16.6	0.75	16.6	0.75	26.9	0.98
Average	17.7	0.77	17.7	0.77	17.7	0.77	24.4	0.95

Because we take into account the sounds generated by Laplacian operation, the suggested approach appears to be a denosing method. These sounds, however, cannot be categorized as Gaussian, Speckle, or Pepper noises [29]. Therefore, it is challenging to create a denoising technique to eliminate the noises generated by Laplacian operation. As a result, we consider the supplied image to be of low quality due to noise. As a consequence, we provide a fresh improvement model in this study that suppresses sounds and produces an improved image. We test the proposed model on three common noisy photos—Lena, Cameraman, and Boat—by adding Gaussian noises with a standard deviation of 15 as described in [29] in order to determine how well it performs on noise images.In Fig.3, the three benchmark photos are displayed (a). In Fig. 3, noisy pictures with a standard deviation of 15 Gaussian noises have been added (b). Due to the loss of information in the final pictures, the suggested model on noisy images in Fig. 5.3(c) does not produce satisfactory results. The same conclusions may be derived from the current enhancement models applied to noisy pictures, as shown in Figs. 3(a) through (c), where it is not evident if the noises were eliminated by the enhancement models or not. In Table 2, which compares the quantitative results of the proposed model with the current enhancement models for noisy pictures, it is clear that the suggested model performs worse than the current enhancement models in terms of PSNR and SSIM.The suggested model was created to eliminate actual sounds caused by Laplacian operation, but not additional noises. The results provided by the existing enhancement models are inconsistent when compared to the findings in Tables 2 and 3, while the suggested model provides the best results for Laplacian noisy pictures and the poorest results for other noisy images.

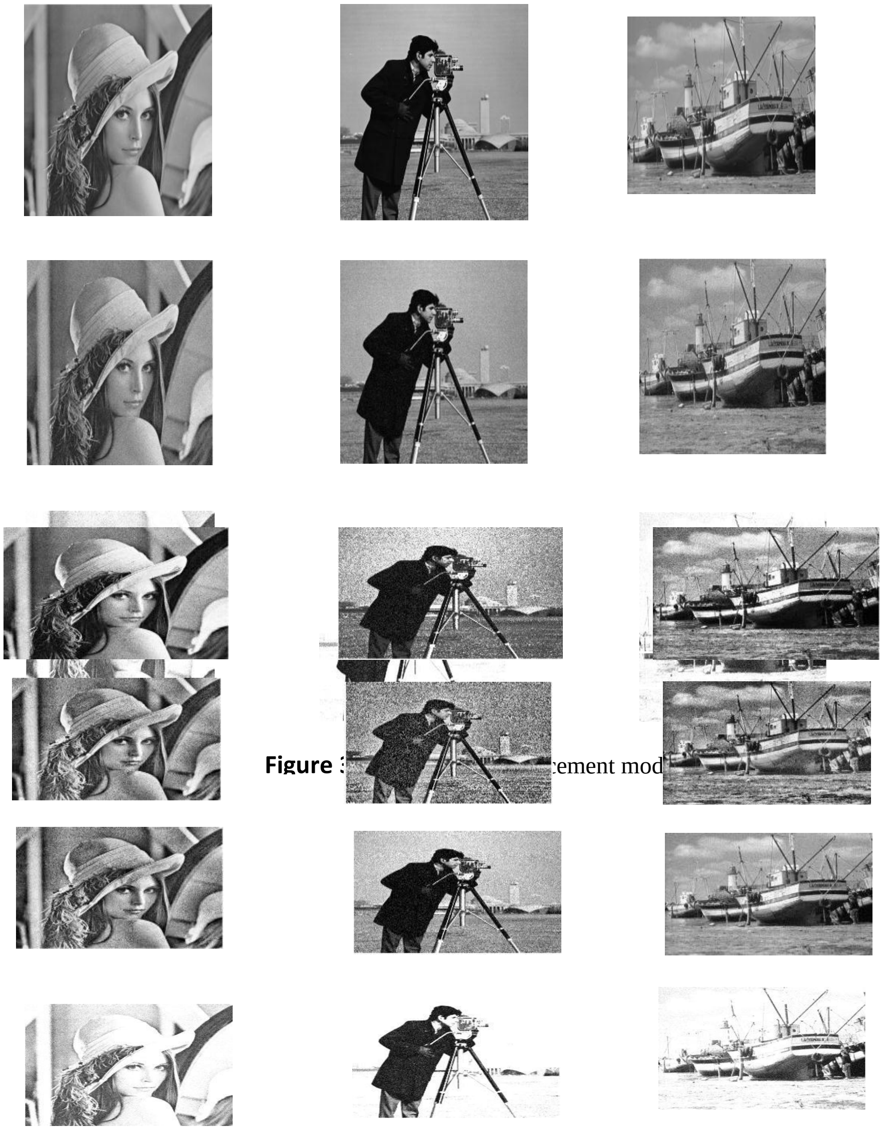


Figure 3 cement mod

Figure 3:(d) Enhanced Results of the

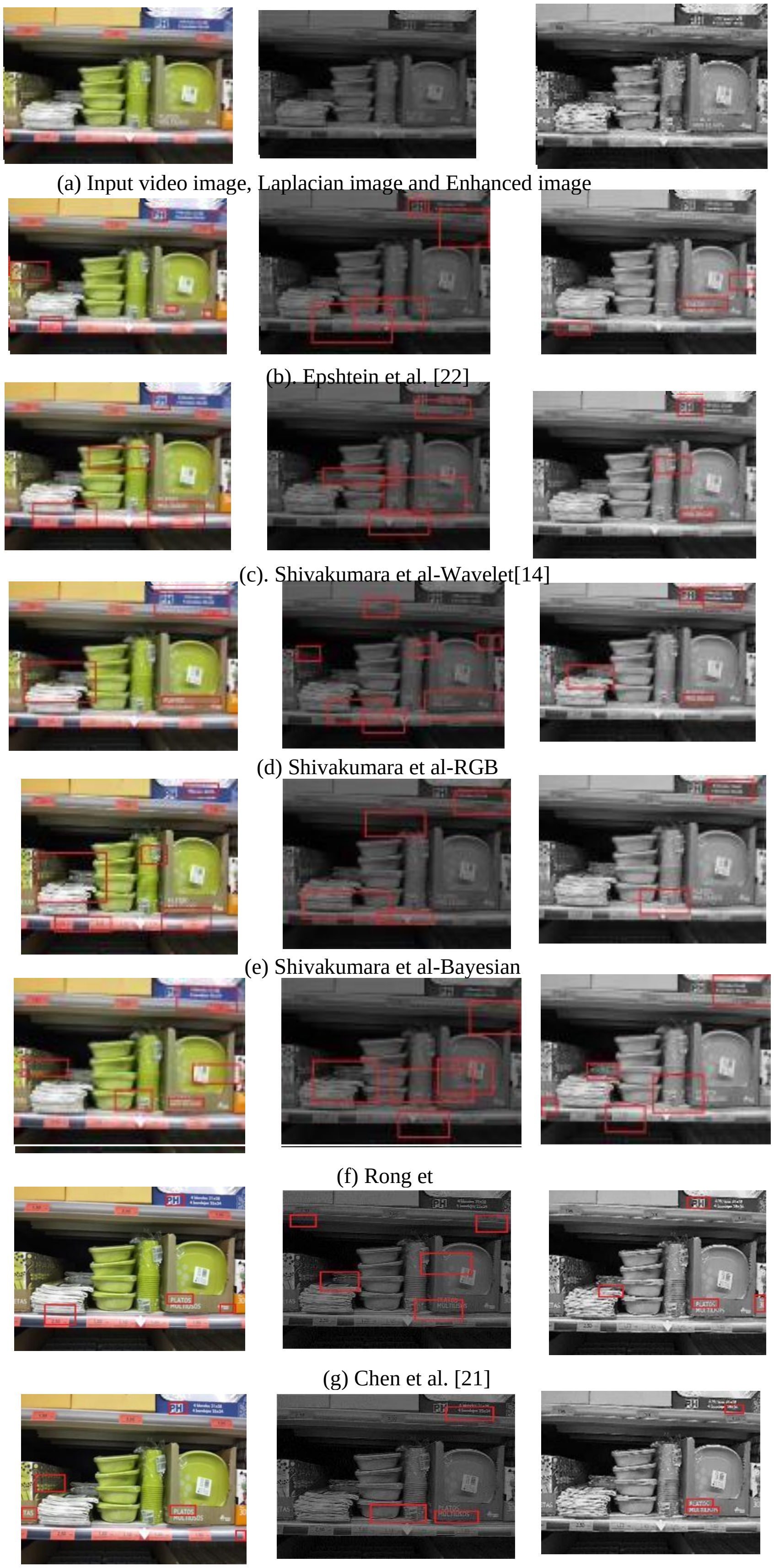
Table 3: Performance of the proposed and existing techniques on standard noisy mages corrupted by Gaussian noise with standard deviation, 15

Images	HE		CLAHE		AIV		Proposed Model	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Lena	17.74	0.62	4.88	0.49	21.33	0.70	9.16	0.59
Cameraman	17.69	0.36	14.30	0.27	18.75	0.45	9.98	0.57
Boat	15.87	0.56	14.15	0.51	20.97	0.77	8.89	0.62
Average	17.10	0.52	11.11	0.42	20.35	0.64	9.34	0.59

B. Validating Enhancement Model through Text Detection

As mentioned in the preceding section, rather than manually introducing noise to the input picture as denoising approaches do in the literature [29], we use the noises produced by Laplacian operation as a case study for creating the enhancement model. Corrupted data therefore lack a basis for determining quality metrics. For the purpose of calculating a quality measure, we use an input image as the original image. We therefore conduct experiments on text detection and recognition of the texts in the input image, the Laplacian image, and the enhanced image to demonstrate the usefulness of the proposed enhancement model. We show that text detection and recognition methods give better accuracies for enhanced images compared to input and Laplacian images.

Figure 5.4 displays an example set of qualitative text detection algorithm results for the ICDAR 2013 video dataset. For the purpose of validating the outcomes of the suggested improvement model, we employ several text detection techniques. Figure 7 illustrates how all text detection techniques produce subpar detection results for the input photos, with the Laplacian images showing the poorest results due to noise introduction but improved results for the upgraded images. Fig. 4 further shows that, in comparison to the input and upgraded photos, text detection algorithms often produce higher false positives for Laplacian images. This suggests that the sounds produced by Laplacian operation have a major negative impact on the effectiveness of text identification techniques as a whole.As a result, we can say that the improved photos greatly increase the text detection techniques' capabilities. The qualitative findings of various datasets, like ICDAR 2013 scene data, SVT data, and MSRA data, can provide the same conclusions. Table 4 to Table 7 present the quantitative results of the text detection methods for the aforementioned datasets. From these results, it can be seen that the text detection methods perform worse on Laplacian images than on input and enhanced images, while performing better on enhanced images than on input and Laplacian images. The sounds introduced by Laplacian operation make this



(a) Input video image, Laplacian image and Enhanced image

(b). Epshtein et al. [22]

(c). Shivakumara et al-Wavelet[14]

(d) Shivakumara et al-RGB

(e) Shivakumara et al-Bayesian

(f) Rong et

(g) Chen et al. [21]

(h) Yin et al. [22]

Figure 4 :The sample qualitative resutls of text detection methods for the ICDAR 2013 video dataset

Table 4: Text detection before and after enhancement on ICDAR 2013 Video data

Methods	Before Enhancement						After Enhancement		
	Original			Laplacian			Enhanced		
	R	P	F	R	P	F	R	P	F
Epshtein et al. [24]	57.23	50.17	53.46	38.15	25.2	30.35	53.25	63.1	57.75
Shivakumara et al-Wavelet [14]	38.12	62.32	47.30	39.46	28.12	32.83	46.23	61.13	52.64
Shivakumara et al-RGB [23]	65.1	42.62	51.51	73.4	26.21	38.62	65.12	52.12	57.89
Shivakumara et al-Bayesian [15]	55.21	29.19	38.18	45.11	21.87	29.45	50.21	34.22	40.70
Rong et al. [20]	36.0	25.12	29.59	30.12	13.10	18.25	45.0	22.5	30.0
Chen et al. [21]	51.16	57.48	54.13	31.41	35.12	33.16	55.83	59.24	57.48
Yin et al. [22]	52.73	64.31	57.94	48.21	52.6	50.30	57.27	64.3	60.58

Table 5: Text detection before and after enhancement on ICDAR 2013 Scene

Methods	Before Enhancement						After Enhancement		
	Original			Laplacian			Enhanced		
	R	P	F	R	P	F	R	P	F
Epshtein et al. [24]	63.0	70.0	66.31	87.5	46.6	60.81	65.03	72.23	68.44
Shivakumara et al-Wavelet [14]	49.0	58.0	53.12	49.5	40.4	44.48	60.21	54.12	57.00
Shivakumara et al-RGB [23]	61.37	41.51	49.52	50.27	36.51	42.29	74.23	41.24	53.02
Shivakumara et al-Bayesian [15]	47.46	29.63	36.48	48.8	16.41	24.56	45.11	33.21	38.25
Rong et al. [20]	47.30	28.0	35.17	40.3	12.0	18.43	47.56	36.0	40.98
Chen et al. [21]	62.26	69.83	65.82	67.78	48.37	56.45	60.57	80.15	69.00
Yin et al. [22]	64.47	77.56	70.41	60.32	68.76	64.26	67.32	78.43	72.45

Table 6: Text detection before and after enhancement on SVT

Methods	Before Enhancement						After Enhancement		
	Original			Laplacian			Enhanced		
	R	P	F	R	P	F	R	P	F
Epshtein et al. [24]	32.88	48.52	39.19	38.21	21.63	27.62	34.12	64.12	44.53
Shivakumara et al-Wavelet [14]	48.1	30.41	37.26	50.6	20.26	28.93	50.47	39.12	44.07

Shivakumara et al-RGB [23]	28.12	15.2	19.73	18.37	10.34	13.23	32.18	22.17	26.25
Shivakumara et al-Bayesian [15]	28.37	20.61	23.87	33.17	13.48	19.16	29.12	21.13	24.48
Rong et al. [20]	32.17	23.46	27.13	47.12	14.6	22.29	47.13	18.5	26.57
Chen et al. [21]	34.15	50.82	40.84	30.7	35.17	32.78	38.35	56.75	45.76
Yin et al. [22]	38.11	50.74	43.52	33.78	47.2	39.37	39.89	59.43	47.73

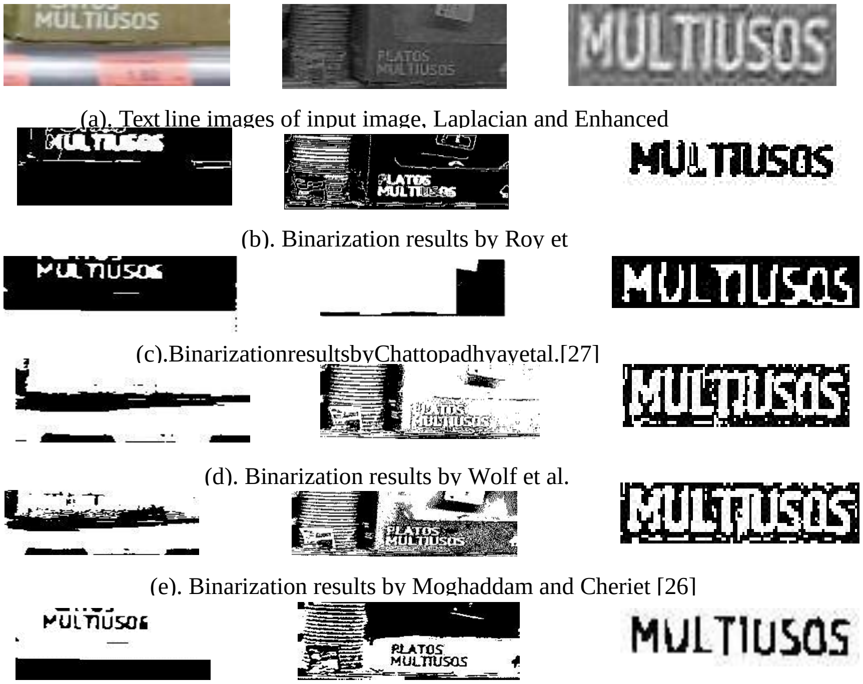
Table 7: Text detection before and after enhancement on MSRA

Methods	Before Enhancement						After Enhancement		
	Original			Laplacian			Enhanced		
	R	P	F	R	P	F	R	P	F
Epshtein et al. [24]	30.65	24.12	26.99	31.2	15.35	20.57	32.5	24.12	27.68
Shivakumara et al-Wavelet [14]	50.12	40.3	44.67	51.12	30.12	37.90	45.32	52.12	48.48
Shivakumara et al-RGB [23]	35.85	15.64	21.77	30.51	12.48	17.71	34.17	21.25	26.20
Shivakumara et al-Bayesian [15]	47.72	31.1	37.65	21.13	27.5	30.22	50.14	38.7	43.68
Rong et al. [20]	41.43	17.0	24.10	28.17	17.45	21.55	36.13	28.12	31.62
Chen et al. [21]	36.48	29.73	32.76	28.39	24.47	26.28	35.26	34.57	34.91
Yin et al. [22]	57.38	63.7	60.37	53.65	55.58	54.59	59.18	64.73	61.83

c. Validating Enhancement Model throughRecognition

This section offers recognition tests for the text lines found by text detection methods on the input, Laplacian, and improved pictures, comparable to text detection studies. As it works for both video and photos of real-world scenes, we utilize the approach in [14] for text line identification from each input image, Laplacian image, and enhanced image for recognition studies. Figure 5 displays the qualitative outcomes of several binarization techniques for various datasets, including the ICDAR 2013 video.In contrast to the outcomes in the improved picture, it can be shown from Fig. 5 that the binarization approaches produce subpar results for the input image and Laplacian image. The same holds true for additional datasets, such as the MSRA, SVT, and ICDAR 2013 scene data. Table 8 reports the quantitative outcomes of the binarization techniques. The binarization approaches produce subpar results on video data when we compare the findings from the ICDAR 2013 video and ICDAR scene datasets. This is due to the fact that video data often has lower resolution than camera-based pictures, such as those seen in ICDAR scene, SVT, and MSRA.Similar to this, when we compare SVT and MSRA data, the approaches produce bad results for MSRA but excellent results for SVT because SVT data has essentially horizontal texts whereas MSRA has texts that are arbitrarily oriented. Even when binarization techniques produce good results for texts with diverse orientations, OCR engines can struggle to accurately identify them because of their inherent flaws. Additionally, for all datasets, relative to the input and improved photos, all binarization techniques produce subpar Laplacian image outcomes.The recognition rates of the improved photos outperform those of the input images when we compare the results of the enhanced and input images. This demonstrates how effective the suggested enhancement approach is in enhancing recognition outcomes.

In conclusion, it can be concluded from the foregoing talks of text detection and recognition tests that, while text detection and recognition systems perform worse for Laplacian pictures, they perform much better for enhanced images.



(a). Text line images of input image, Laplacian and Enhanced

(b). Binarization results by Roy et

(c).BinarizationresultsbyChattopadhyayetal.[27]

(d). Binarization results by Wolf et al.

(e). Binarization results by Moghaddam and Cheriet [26]

"MULTIUSOS"

(f). Binarization results by Howe [18]

Figure 5 : displays Binarization results

Table 8: Recognition results Before Enhancement (BE) and After Enhancement (AE)

(Note: Shivakumara et al-wavelet [14] is used for text line detection from both original, Laplacian and denoising image. Here RR: Recognition Rate, O: Original Image, L: Laplacian Image, E: Enhanced Image)

Methods	ICDAR 2013 Video			ICDAR 2013 Scene			SVT			MSRA		
	BE		AE	BE		AE	BE		AE	BE		AE
	O	L	E	O	L	E	O	L	E	O	L	E
	RR	RR	RR	RR	RR	RR	RR	RR	RR	RR	RR	RR
Roy et al. [25]	21.12	16.12	35.7	24.12	21.6	48.12	21.32	14.12	37.1	16.5	11.7	18.12
Chattopadhyay et al. [27]	23.68	23.8	36.12	21.68	22.68	47.76	23.18	12.21	34.21	17.1	10.8	19.74
Wolf et al.[28]	24.1	18.1	32.51	25.1	19.1	42.83	23.6	15.12	36.24	16.5	11.4	18.8
Moghaddam and Cheriet [26]	27.1	24.7	39.34	22.1	20.52	50.12	22.5	18.27	41.1	15.12	11.2	17.23
Howe [18]	12.2	26.5	41.2	29.8	23.2	53.35	27.2	17.57	43.12	18.6	14.5	21.41

5. Conclusion

By taking into account the sounds produced by Laplacian operation, we have presented a novel Fractional Poission model for improving fine details in video pictures as well as natural scene photographs. We believe that the suggested model also performs well for additional distortions and noises that are manually introduced because it performs well for the noises produced by the Laplacian operation. For each of the four benchmark datasets, the quality metrics of the suggested model produce better results than those of the baseline augmentation strategies. We run text detection and identification tests using various input methods, such as Laplacian and enhanced pictures, to demonstrate the utility of the proposed model.The results of the experimental tests demonstrate that the suggested model considerably enhances the performance of text detection and recognition algorithms in terms of several metrics. In contrast to scanned papers from the field of document analysis, the text detection and recognition accuracy reported in this work is still lower. Our future work will concentrate on further developing the enhancement model so that the text detection and identification techniques now in use can achieve accuracy levels comparable to those found in the field of document analysis by investigating temporal information in the case of video.

REFERENCES

[1] A. Balagopalan, L. L. Balasubramanian, V. Balasubramanian, N. Chandrasekharan and A. Damodar, "Automatic keyphrase extraction and segmentation of video lectures," 2012 IEEE International Conference on Technology Enhanced Education (ICTEE), Amritapuri, India, 2012, pp. 1-10, doi: 10.1109/ICTEE.2012.6208622.

[2] A. Kumar and N. Awasthi, "An efficient algorithm for text localization and extraction in complex video text images," 2013 2nd International Conference on Information Management in the Knowledge Economy, Chandigarh, India, 2013, pp.14-19.

[3] Aloysius George and B. R. Rajakumar, "APOGA: An Adaptive Population Pool Size based Genetic Algorithm", AASRI Procedia - 2013 AASRI Conference on Intelligent Systems and Control (ISC 2013), Vol. 4, pages: 288-296, 2013, DOI: https://doi.org/10.1016/j.aasri.2013.10.043.

[4] Alsattar, H.A., Zaidan, A.A. & Zaidan, B.B. Novel meta-heuristic bald eagle search optimisation algorithm. Artif Intell Rev 53, 2237–2264 (2020). https://doi.org/10.1007/s10462-019-09732-5.

[5] Amolkumar Narayan Jadhav,Gomathi N, "DIGWO: Hybridization of Dragonfly Algorithm with Improved Grey Wolf Optimization Algorithm for Data Clustering", Multimedia Research, Vol.2,No.3, pp.1-11,2019.

[6] Aneeshan SainAyan Kumar BhuniaUmapada Pal, "Multi-oriented text detection and verification in video frames and scene images", Neurocomputing, Volume 275 (Cover date: 31 January 2018), Pages 1531-1549, 4 October 2017.

[7] Anshul MittalPartha Pratim RoyBalasubramanian Raman, "Rotation and script independent text detection from video frames using sub pixel mapping", Journal of Visual Communication and Image Representation, Volume 46, Pages 187-198, July 2017.

[8] Anurag AgrahariRajib Ghosh, "Multi-Oriented Text Detection in Natural Scene Images Based on the Intersection of MSER With the Locally Binarized Image", Procedia Computer Science, Volume 171 (Cover date: 2020), Pages 322-330, 4 June 2020.

[9] B. Bouaziz, T. Zlitni and W. Mahdi, "AViTExt: Automatic Video Text Extraction; A new Approach for video content indexing Application," 2008 3rd International Conference on Information and Communication Technologies: From Theory to Applications, Damascus, Syria, 2008, pp. 1-5, doi: 10.1109/ICTTA.2008.4530004.

[10] B.Epshtein,E.Ofek,Y.Wexler(2010).Detectingtextinnaturalsceneswithstrokewidth transform. In: Proc. CVPR,2963-2970.

[11] B. R. Rajakumar and Aloysius George, "A New Adaptive Mutation Technique for Genetic Algorithm", In proceedings of IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pages: 1-7, December 18-20, Coimbatore, India, 2012, DOI: 10.1109/ICCIC.2012.6510293.

[12] B. R. Rajakumar, "Impact of Static and Adaptive Mutation Techniques on Genetic Algorithm", International Journal of Hybrid Intelligent Systems, Vol. 10, No. 1, pages: 11-22, 2013, DOI: 10.3233/HIS-120161.

[13] B. R. Rajakumar, "Static and Adaptive Mutation Techniques for Genetic algorithm: A Systematic Comparative Analysis", International Journal of Computational Science and Engineering, Vol. 8, No. 2, pages: 180-193, 2013, DOI: 10.1504/IJCSE.2013.053087.

[14] Bernt Schiele and James L. Crowley. Object recognition using multidimensional receptive field histograms. In Proc. European Conf. Computer Vision, pages 610–619, Cambridge, UK, 1996.

[15] C.Chen,L.Zhang,J.Bu,C.WangandW.Chen(2010).ConstrainedLaplacianEigenmap for dimenionality reduction. Neurocomputing,951-958.

[16] C.M.Gracia,M.MirmehidandJ.L.G.Mora(2013).Fastpersepectiverecoveryoftextin natural scenes. Image and Vision Computing,714-724.

[17] C. Shi, Q. Ruan and G. An (2014). Sparse feature selection on graph Laplacian for web image annotation. IVC,189-201.

[18] C. Wolf, J. Michel. Jolion and F. Chassaing (2002). Text Localization, Enhancement and Binarization in Multimedia Documents. In Proc. ICPR,1037-1040.

[19] C. Yao1, X. Bai1, W. Liu, Y. Ma and Z. Tu (2012). Detecting Texts of Arbitrary Orientations in Natural Images. In Proc. CVPR,1083-1090.

[20] C. Yi and Y. Tian (2013). Text extraction from scene images by character apperance and structure modeling. CVIU, 2013,182-194.

[21] C. Zhou, T. Yan, W. Tao and S. Lui (2012). A Study of Images Denoising Based onTwo Improved Fractional Integral Marks. ICT, LNCS, 386-392.

[22] Chang LiuChun YangXu-Cheng Yin, "GCCNet: Grouped channel composition network for scene text detection", Neurocomputing, Volume 454 (Cover date: 24 September 2021), Pages 135-151, 5 May 2021.

[23] Chaudhuri, A., Shivakumara, P., Chowdhury, P.N. et al. A deep action-oriented video image classification system for text detection and recognition. SN Appl. Sci. 3, 838 (2021). https://doi.org/10.1007/s42452-021-04821-z.

[24] Chen, Datong & Odobez, Jean-Marc & Bourlard, Herve. (2004). Text Detection and Recognition in Images and Videos.

[25] Cheng et al., "FREE: A Fast and Robust End-to-End Video Text Spotter," in IEEE Transactions on Image Processing, vol. 30, pp. 822-837, 2021, doi: 10.1109/TIP.2020.3038520.

[26] Chun Mao & Xiao Hu (2018) Text information extraction algorithm of video images in multimedia environment, Journal of Discrete Mathematical Sciences and Cryptography, 21:2, 305-310, DOI: 10.1080/09720529.2018.1449304.

[27] D.Karatzas,F.Shafait,S.Uchida,M.Iwamura,L.G.I.Boorda,S.R.Mestre,J.Mas,D.F. Mota, J. A. Almazan and L. P. De las Heras (2013). ICDAR 2013 robust reading competition. In Proc. ICDAR,1115-1124.

[28] D. Rajendran, P. Shivakumara, B. Su, S. Lu and C. L. Tan, "A New Fourier-Moments Based Video Word and Character Extraction Method for Recognition," 2011 International Conference on Document Analysis and Recognition, Beijing, China, 2011, pp. 1165-1169, doi:10.1109/ICDAR.2011.235.

[29] E. Feig and S. Winograd. Fast algorithms for the discrete cosine transform. IEEE Trans. Signal Processing, 40(28):2174–2193, Sept. 1992.

[30] Elshahaby, H., Rashwan, M. A system for detection of moving caption text in videos: a news use case. Multimed Tools Appl 80, 25607–25631 (2021). https://doi.org/10.1007/s11042-021-10856-6.Z.

[31] H. A. Jalab and R. W. Ibrahim (2015). Fractional Alexander polynomials for image denoising. Signal Processing, 340–354.

[32] H. Chen, S. S. Tsai, G. Schorth, D. M. Chen, R. Grzeszczuk and B. Girod (2011).Robust text detection in natural scene images with edge-enhaned maximally stable extremal regions. In Proc. ICIP, pp2609-2612.

[33] H. Shih and C. Huang, "Content Extraction and Interpretation of Superimposed Captions for Broadcasted Sports Videos," in IEEE Transactions on Broadcasting, vol. 54, no. 3, pp. 333-346, Sept. 2008, doi: 10.1109/TBC.2008.2001143.

[34] H. Shih and C. Huang, "Content Extraction and Interpretation of Superimposed Captions for Broadcasted Sports Videos," in IEEE Transactions on Broadcasting, vol. 54, no. 3, pp. 333-346, Sept. 2008, doi: 10.1109/TBC.2008.2001143.

[35] https://in.mathworks.com/help/vision/ug/automatically-detect-and-recognize-text-in-natural-images.html

[36] https://rrc.cvc.uab.es/?ch=3&com=downloads

[37] https://www.techtarget.com/searchcontentmanagement/definition/OCR-optical-character-recognition

[38] Huiwen Cai, Xiaoyan Wang, Ming Xia, Yangsheng Wang, "Entropy-Based Maximally Stable Extremal Regions for Robust Feature Detection", Mathematical Problems in Engineering, vol. 2012, Article ID 857210, 7 pages, 2012. https://doi.org/10.1155/2012/857210

[39] I. Podlubny (1999). Fractional Differential Equations. Academic, New York,1999.

[40] J. Galvis, S. Morales, C. Kasmi and F. Vega, "Denoising of Video Frames Resulting From Video Interface Leakage Using Deep Learning for Efficient Optical Character Recognition," in IEEE Letters on Electromagnetic Compatibility Practice and Applications, vol. 3, no. 2, pp. 82-86, June 2021, doi: 10.1109/LEMCPA.2021.3073663.

[41] J. Toriwaki and S. Yokoi. Distance transformations and skeletons of digitized pictures with applications. In L. N. Kanal and A. Rosenfeld, editors, Progress in Pattern Recognition, pages 187–264. North–Holland, Amsterdam, 1981.

[42] J. Zhang and R. Kasturi (2014). A Novel Text Detection System based on Character and Link Energies. IEEE Trans. IP,4187-4198.