Available online www.ejaet.com

European Journal of Advances in Engineering and Technology, 2017, 4(12): 896-901

Research Article ISSN: 2394 - 658X

Fuzzy Convolutional Neural Network Based Noise Reduction Technique in Scene Text Detection and Recognition

Vaibhav Kumar and ML Garg

Department of Computer Science and Engineering, DIT University, Dehradun, India
vaibhav05cse@gmail.com

ABSTRACT

This paper proposes a scheme for noise reduction in low illumination images. There are various methods already in existence for noise reduction but hybrid models have given better results many times. The proposed scheme is based on a Hybrid Neuro-Fuzzy Model. This model may be termed as Fuzzy Convolutional Neural Network. While detecting and recognizing texts present in the scene images, there is an issue of noise. This noise may be present due to low intensity of light. Since the proposed model will be an integration of Neural Network and Fuzzy Logic, hence it will have the capability to learn and to handle uncertainties present in the noise affected image. This model will reduce the noise present in the image and will present a better recognition of texts in scene images.

Keywords: Scene Text Detection and Recognition, Convolutional Neural Network, Neuro-Fuzzy Systems, Noise Reduction

INTRODUCTION

In the field of image processing, detection and recognition of texts in images and videos has an important role for the vision of an intelligent agent or to develop a system for visually disabled persons [1]. There are several issues associated with the use of various techniques available for this purpose. Images or videos that are captured in the low intensity light shall be an issue for text detection and recognition due to noise [1]. There are several filtering and noise reduction techniques already in existence for images. There are various approaches based on neural networks are used for reduction of noise in images. Neural networks have advantage that they can learn from data but they have limitation that they cannot handle the imprecise information of uncertain data. Fuzzy logic based techniques have advantage that they can handle imprecise information but they have limitation that they cannot learn from the data. The proposed research will integrate fuzzy logic technique with a convolutional neural network to develop such a system that can handle imprecise information or uncertain data and that will also have the capability to learn.

Detection and recognition of texts in images or videos is a challenging problem in computer vision [2]. In recent years it has attracted the attention of researchers to develop techniques for this problem [2]. This attention has increased since there are lot of camera mobile phone available today which can convert the texts into any language [2]. If these mobile phones capture the image and translate the texts in the image [3-4] then it will be a very revolutionary technique since one can know the meaning of any word written in any language. This technique will make anyone be able to know the meaning of a word anywhere, anytime [5].

Although the recognition of text gives rise to many applications, the fundamental goal is to determine whether or not there is text in a given image, and if there is, to detect, localize, and recognize it [6]. In the literature, various stages of these fundamental tasks are referred to by different names including text localization [7], which aims to determine the image positions of candidate text, text detection, which determines whether or not there is text using localization and verification procedures, and text information extraction [8-9], which focuses on both localization and binarization. Tasks such as text enhancement are used to rectify distorted text or improve resolution prior to recognition. Other references include scene text recognition [10] and text recognition in the wild [11], which restrict analysis of images to text in natural scenes. Suffice it to say that the primary goals of text detection, localization and recognition are essential for an ‘end-to-end’ system.

Early text detection and recognition research was a natural extension of document analysis and recognition research, moving from scanned page images to camera captured imagery, focusing on basic pre-processing, detection and

OCR technology [12]. Recently, the application of sophisticated computer vision and learning methods has resulted from the realization that the problems do not lend themselves to a sequential series of independent solutions. The trend is to integrate the detection and recognition tasks into an 'end-to-end' text recognition system [13]. In the early years, researchers extensively investigated graphic overlay text in video as a way to index video content. Scene text, especially video scene text, has been regarded as presenting a more difficult challenge yet very little work had been done with it [14]. Recently, researchers have explored approaches that prove effective for text captured in various configurations, in particular, incidental text in complex backgrounds. Such approaches typically stem from advanced machine learning and optimization methods, including unsupervised feature learning [15], convolutional neural networks (CNN) [11], deformable part-based models (DPMs), belief propagation and conditional random fields (CRF) [16].

CHALLENGES IN SCENE TEXT DETECTION AND RECOGNITION

The complexity of environments, flexible image acquisition styles and variation of text contents pose various challenges.

Scene Complexity: In natural environments, numerous man-made objects, such as buildings, symbols and paintings appear, that have similar structures and appearances to text. Text itself is typically laid out to facilitate legibility. The challenge with scene complexity is that the surrounding scene makes it difficult to discriminate text from non-text.

Uneven Lighting: When capturing images in the wild, uneven lighting is common due to the illumination and the uneven response of sensory devices. Uneven lighting introduces color distortion and deterioration of visual features, and consequently introduces false detection, segmentation and recognition results.

Blurring and Degradation: With flexible working conditions and focus-free cameras, defocusing and blurring of text images occur [17]. Image/video compression and decompression procedures also degrade the quality of text, in particular, graphical video text. The typical influence of defocusing, blurring and degradation is that they reduce characters' sharpness and introduce touching characters, which makes basic tasks such as segmentation difficult [17].

Aspect Ratios: Text such as traffic signs, may be brief, while other text, such as video captions, may be much longer. In other words, text has different aspect ratios. To detect text, a search procedure with respect to location, scale and length needs to be considered, which introduces high computational complexity.

Distortion*:* Perspective distortion occurs when the optical axis of the camera is not perpendicular to the text plane. Text boundaries lose rectangular shapes and characters distort, decreasing the performance of recognition models trained on undistorted samples. Fonts. Characters of italic and script fonts might overlap each other, making it difficult to perform segmentation. Characters of various fonts have large within-class variations and form many pattern sub-spaces, making it difficult to perform accurate recognition when the character class number is large.

Multilingual Environments: Although most of the Latin languages have tens of characters, languages such as Chinese, Japanese and Korean (CJK), have thousands of character classes. Arabic has connected characters, which change shape according to context. Hindi combines alphabetic letters into thousands of shapes that represent syllables. In multilingual environments, OCR in scanned documents remains a research problem [18], while text recognition in complex imagery is more difficult.

Uneven Lighting and Low Illumination Images

Digital imaging devices such as digital cameras or camera-phones are becoming very popular and ubiquitous. More and more people are capturing many digital photographs and the number of digital photographs taken is increasing rapidly. The overall quality of images taken by many of these imagers, however, is not always satisfactory. Many digital cameras or camera phones today provide poor quality shots in low illuminations such as indoors or night situations. It is very difficult to capture good quality photographs in these situations since elongating exposure time increases motion blur and shortening exposure time reduces the signal-to-noise-ratio (SNR). Image sensors in digital cameras typically lack the dynamic range and sensitivity to capture both the dark and bright parts of the scene. In many cases, auto exposure algorithm sets the exposure time such that the bright region is not overly saturated, leaving the dark region grossly under-exposed. Such pictures have very poor signal-to-noise-ratio (SNR) due to lack of captured photons. There are other scenarios, when even carefully planned shots are underexposed, such as in museums where flash is not permitted, or when taking a picture of a moving object at high ISO settings. In addition, low quality optics and sensors are included in many consumer devices, such as camera phones and PDAs. These devices are typically used to take unplanned casual shots, potentially in bad lighting conditions that induce increased noise. Furthermore, many of these devices do not have built-in-flash, making it nearly impossible to capture good quality shots in low light situations [19]. Hence in the images that are captured in low light intensity may have noise and due to this presence of noise detection and recognition of texts in images will not be performed effectively.

NEURAL NETWORKS IN IMAGE PROCESSING

There are more than 200 applications of neural networks in image processing. In which feed-forward Kohonen feature maps, Hopfield neural networks are specially used into 2-dimensional taxonomy [20]. Neural Networks have already been used to reduce noises in images [21]. Convolutional Neural Networks (CNNs) have been mainly used for this purpose [22]. CNNs have shown very effective results in hand-written digits and traffic sign recognition [23] and image denoising [24]. A CNN accepts an image as input and produces output through the layers of convolution and subsampling. Advantage of using CNN in place of Multilayer Perceptron (MLP) is that while training a MLP with many hidden layers can lead to the problem of over fitting and vanishing gradients [25].

Text detection and recognition in natural scene images has applications in computer vision systems such as image retrieval, automatic license plate recognition, automatic street sign translation or help for visually impaired people [5]. For automatic street sign translation systems, text recognition systems recognize text, which is processed by a machine translation tool to translate text into another language. By recognizing license plates automatic toll collection is possible. For visually impaired people text recognition systems can help to recognize and read text on street signs with text-to-speech systems. There are various issues associated with the technique of scene text detection and recognition. One important issues are the images that are captured in low intensity light that is low illumination images. In low illumination images, processing or text recognition is affected due to the presence of noise. To reduce this noise several techniques are used. Neural Network based techniques are used very well in all phases of image processing due to the learning capability of neural nets. But they have limitation that they cannot handle imprecise information. Many uncertainties remain present in noise affected image data. So neural networks may have limitation to be used there. Fuzzy logic based technique can handle imprecise information but the system based on it cannot learn since fuzzy logic based techniques do not have learning capability.

Fuzzy Image Processing

Fuzzy techniques offer a new and flexible framework for the development of image enhancement algorithms. They are nonlinear, knowledge-based and robust. The potentials of fuzzy set theory for image enhancement are still not investigated in comparison with other established methodologies [27]. Because of the ability to handle and manage the imprecision encountered with images effectively, applying fuzzy set theory becomes a strong in road image processing areas like image enhancement [26]. Many research works are still going on in this area to make improvements in the existing techniques. Use of intensification operator makes a boom in the field of enhancement of images. Different algorithms use different membership functions like triangular, Gaussian, exponential, triangular, S-function, or trapezoidal to map pixel image to fuzzy image. An image contrast enhancement algorithm was developed using some trigonometric membership function.

Many research works have been conducted in this field but some problems are not considered. These can be listed out as: -

- Although, many researchers have developed techniques for scene text detection and recognition but low illumination images that is the images captured in low intensity light are still an issue for scene text detection and recognition.
- There is heavy noise in low illumination images. Some better quality noise reduction technique required to be developed.
- Neural Networks have been very successfully applied in image processing but have limitation that they cannot handle imprecise information or incomplete data.
- Fuzzy logic based techniques are successfully applied in the field of image processing [26] but they have limitation that they cannot lean from the data.

Text detection and recognition in natural scene images has applications in computer vision systems such as image retrieval, automatic license plate recognition, automatic street sign translation or help for visually impaired people [5]. For automatic street sign translation systems, text recognition systems recognize text, which is processed by a machine translation tool to translate text into another language. By recognizing license plates automatic toll collection is possible. For visually impaired people text recognition systems can help to recognize and read text on street signs with text-to-speech systems. There are various issues associated with the technique of scene text detection and recognition. One important issues are the images that are captured in low intensity light that is low illumination images. In low illumination images, processing or text recognition is affected due to the presence of noise. To reduce this noise several techniques are used. Neural Network based techniques are used very well in all phases of image processing due to the learning capability of neural nets. But they have limitation that they cannot handle imprecise information. Many uncertainties remain present in noise affected image data. So neural networks may have limitation to be used there. Fuzzy logic based technique can handle imprecise information but the system based on it cannot learn since fuzzy logic based techniques do not have learning capability.

Hybrid Neuro-Fuzzy Model

A hybrid Neuro-Fuzzy system can be developed that have the capability to handle incomplete, imprecise, vague data and also have the capability to learn from the data. Convolutional Neural Networks are used very well in image processing but due to uncertainties present in noise affected image, they have limitation to be used in denoising. If Fuzzy Logic and Convolutional Neural Network are integrated together then a very efficient noise reduction technique may be developed. A Fuzzy Convolutional Neural Network based technique is proposed to be developed to reduce noise of low illumination images in scene text detection and recognition which may avoid the issue of images that are captured in low intensity light while detecting and recognizing texts in scene or natural images and videos.

The main objective of this research is to develop a Fuzzy Convolutional Neural Network based technique for reducing the noise in low illumination images so as to effectively detect and recognize texts in scene images. This approach is expected to produce better results in its performance as compared to other existing techniques.

RESEARCH METHODOLOGY

For developing a hybrid Neuro-Fuzzy system – Fuzzy Convolutional Neural Network- the following methods will be followed: -

Fuzzy Image Processing

Fuzzy image processing is not a unique theory. It is a collection of different fuzzy approaches to image processing [28]. Nevertheless, the following definition can be regarded as an attempt to determine the boundaries:

Fuzzy image processing is the collection of all approaches that understand, represent and process the images, their segments and features as fuzzy sets. The representation and processing depend on the selected fuzzy technique and on the problem to be solved [28].

Fuzzy image processing has three main stages: image fuzzification, modification of membership values, and, if necessary, image defuzzification.

Convolutional Neural Network

Convolutional neural networks are inspired by biological process of visualization. They are the variants of multilayer perceptron. They are designed to use minimum amount of pre-processing [29]. They have a very important application in image processing. It is represented in figure 1. Due to the weight sharing architecture of convolutional neural network, learning can be performed with a modification in backpropagation algorithm [30]. The image denoising task may be framed as learning for this network [21].

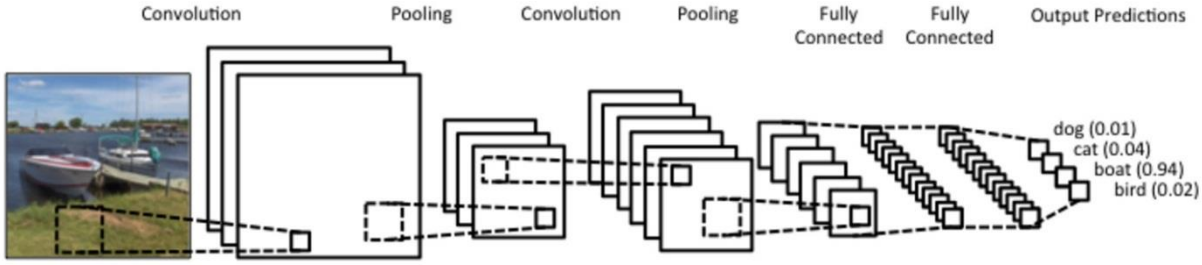

Fig. 1 Convolutional Neural Network

Fuzzy Convolutional Neural Network

A Fuzzy Convolutional Neural Network is proposed to be developed in this research that can be used for noise reduction of low illumination images in scene text detection and recognition. This system will be developed by integrating fuzzy logic techniques with a convolutional neural network. This system will have the capability of learning and to handle imprecise information.

PERFORMANCE EVALUATION

Performance of various methods for noise reduction in images can be compared on the basis of following parameters: -

- **Mean**
- **Standard Deviation**
- **Mean Squared Error (MSE**)
- **Peak Signal-to-Noise Ratio (PSNR)**- PSNR represents a measure of the peak error and it can be calculated from above three parameters.

PSNR is most easily defined via the mean squared error (*MSE*). Given a noise-free $m \times n$ monochrome image f and its noisy approximation *g*, *MSE* is defined as:

$$\text{MSE} = \frac{1}{mn}\sum_{0}^{m-1}\sum_{0}^{n-1}\|f(i,j) - g(i,j)\|^2 \quad (1)$$

The PSNR (in dB) is defined as:

$$PSNR = 20\log_{10}\left(\frac{MAX_f}{\sqrt{MSE}}\right) \quad (2)$$

Here, MAX_I is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, this is 255. More generally, when samples are represented using linear PCM with B bits per sample, MAX_I is 2^B-1. Here MSE and PSNR are the error matrices which represent cumulative squared error and measure of the peak error respectively between original image and the resultant image.

CONCLUSION

An efficient method for noise reduction in low illumination images is proposed to be developed in this paper which is useful to resolve the issue of low intensity light in scene text detection and recognition. There are many possible applications of detection and recognition of texts in scene imageries and videos. After improving the recognition using this proposed research, this will create an opportunity for developing the applications for text recognition in indoor or poor light conditions. The proposed method is expected to provide better results as compared to the existing techniques available in the literature which shall be authenticated trough the comparison of their performance on a standard data set.

REFERENCES

[1] Qixiang Ye and David Doermann, Text Detection and Recognition in Imagery: A Survey, *IEEE Transactions On Pattern Analysis and Machine Intelligence*, **2015**, 37 (7), 1480-1500.
[2] Gökhan Yildirim, Radhakrishna Achanta and Sabine Süsstrunk, Text Detection and Recognition in Natural Images, *8th International Conference on Computer Vision Theory and Applications*, Barcelona, Spain, **2013**.
[3] C Liu, C Wang and R Dai, Text Detection in Images based on Unsupervised Classification of Edge-Based Features, Eighth International Conference on Document Analysis and Recognition, Seol, South Korea, **2005**, 610–614.
[4] X Liu and D Doermann, A Camera Phone Based Currency Reader for the Visually Impaired, The International *ACM SIGACCESS Conference* on *Computers* and Accessibility, Nova Scotia, Canada, **2008**, 305–306.
[5] Michael Opitz, *Text Detection and Recognition in Natural Scene Images*, A Technical Report at Computer Vision Lab, Institute of Computer Aided Automation, Vienna University of Technology, **2013**.
[6] M. Nachtegael et al, The Possibilities of Fuzzy Logic in Image Processing, *International Conference on Pattern Recognition and Machine Intelligence*, Berlin Heidelberg, 2007, 198-208.
[7] R Lienhart and A Wernicke, Localizing and segmenting text in images and videos, *IEEE Transaction on Circuits System for Video Technology*, **2002**, 12 (4), 256–268.
[8] K Jung, KI Kim and AK Jain, Text Information Extraction in Images and Video: A Survey, *Pattern Recognition*, **2004**, 37, 977–997.
[9] J Zang and R Kasturi, Extraction of Text Objects in Video Documents: Recent Progress, *Proceedings of IAPR International Workshop on Document Analysis System*, Nara, Japan, **2008**, 5–17.
[10] JJ Weinman, E Learned-Miller and A Hanson, Scene Text Recognition using Similarity and a Lexicon with Sparse Belief Propagation, *IEEE Transactions* on *Pattern Analysis* and *Machine Intelligence*, **2009**, 31 (10), 1733–1746.
[11] T Wang, DJ Wu, A Coates and AY Ng, End-to-end Text Recognition with Convolution Neural Networks, Proc. *IEEE International Conference on Pattern Recognition*, Tsukuba, Japan, **2012**, 3304–3308.
[12] A Vinciarelli, A Survey on Off-Line Word Recognition, *Pattern Recognition*, **2002**, 35 (7), 1433–1446.
[13] K Wang and S Belongie, Word Spotting in the Wild, *Proceedings of European Conference on Computer Vision*, **2010**, 591–604.
[14] A Coates, B Carpenter, C Case, S Satheesh, B Suresh, T Wang, DJ Wu and AY Ng, Text Detection and Character Recognition in Scene Images with Unsupervised Feature Learning, *Proceedings of IEEE International Conference on Document Analysis and Recognition*, **2011**, 440–445.
[15] C Liu, C Yang, XQ Ding and K Wang, An Improved Scene Text Extraction Method using Conditional Random Field and Optical Character Recognition, *Proceedings of IEEE International Conference on Document Analysis and Recognition*, **2011**, 708–712.
[16] J Liang, D Doermann and H Li, Camera-based Analysis of Text and Documents: A Survey, *International Journal of Document Analysis and Recognition*, **2005**, 7, 84–104.
[17] R Smith, D Antonova and D Lee, Adapting the Tesseract Open Source OCR Engine for Multilingual OCR, *Proceedings of Joint Workshop on Multilingual OCR Analysis of Noisy Unstructured Text Data*, **2011**.
[18] Suk Hwan Lim, Characterization of Noise in Digital Photographs for Image Processing, *IS&T/SPIE Electronic Imaging*, 2008, Vol. 6069.

[19] M Egmont-Petersen, D de Ridder and H Handels, Image Processing with Neural Networks—A Review, *Pattern Recognition*, **2002**, 35(10), 2279-2301.
[20] V Jain and H Seung. Natural Image Denoising with Convolutional Networks, *Advances in Neural Information Processing Systems* (NIPS), **2008**, 21, 769–776.
[21] Y Le Cun, L Bottou, Y Bengio, P Haffner, Gradient-Based Learning Applied to Document Recognition, Proceedings of the IEEE (*Conference name not given in the paper*), **1998**, 86(11), 2278–2324.
[22] P Sermanet and Y Le Cun, Traffic Sign Recognition with Multi-Scale Convolutional Networks, Proceedings of *International Joint Conference on Neural Networks*, CA, USA, **2011.**
[23] Viren Jain, Sebastian Seung, Natural Image Denoising with Convolutional Networks, *Neural Information Processing Systems Conference Proceedings*, **2008.**
[24] G Hinton, S Osindero and Y The, A Fast Learning Algorithm for Deep Belief Nets, *Neural Computation*, **2006**, 18(7), 1527–1554.
[25] SK Pal and RA King, Image Enhancement Using Fuzzy Sets, *IEEE Electronics Letters*, **1980**, 16 (10), 376-378.
[26] VL Jaya and R Gopikakumari, Fuzzy Rule Based Enhancement in the SMRT Domain for Low Contrast Images, *Procedia Computer Science, Elsevier*, 2015, 46, 1747-1753.
[27] Hamid R Tizhoosh, *Fuzzy Image Processing*, Springer Publication, **1997**.
[28] Masakazu Matusugu, Katsuhiko Mori, Yusuke Mitari, Yuji Kaneda, subject Independent Facial Expression Recognition with Robust Face Detection using a Convolutional Neural Networks , *Neural Networks*, **2013**, 16 [5-6], 555-559.
[29] Y Le Cun, B Boser, JS Denker, D Henderson, RE Howard, W Hubbard and LD Jackel, Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, **1989**, 1(4), 4, 541-551.