# TEXT EXTRACTION FROM LOW QUALITY IMAGES USING DEEP LEARNING TECHNIQUES

**PROJECT REPORT**

***Submitted by***

**ASHA SIMON**

**REG NO : TKM20CSCE02**

***In partial fulfillment for the award of the degree of***

**MASTER OF TECHNOLOGY**
**IN**
**COMPUTER SCIENCE AND ENGINEERING**

**Under the guidance of**
**Dr. Aneesh G Nath**


**Thangal Kunju Musaliar College of Engineering**
**Kerala**

**SEPTEMBER 2022**

**Thangal Kunju Musaliar College of Engineering**
**Dept. of Computer Science & Engineering**


# C E R T I F I C A T E

This is to certify that this report titled ***Text Extraction From Low Quality Images Using Deep Learning Techniques*** is a bonafide record of the **Project** presented by **ASHA SIMON (TKM20CSCE02)**, under our guidance and supervision, in partial fulfillment of the requirements for the award of the degree, **M.Tech in Computer Science & Engineering** in **APJ Abdul Kalam Technological University** .

| Coordinator | Supervisor | Head of the Department |
|---|---|---|
| Dr. Ansamma John | Dr. Aneesh G Nath | Dr. Dimple A Shajahan |
| Professor | Associate Professor | Associate Professor |
| Dept. of CSE | Dept. of CSE | Dept. of CSE |
| TKMCE | TKMCE | TKMCE |

# ACKNOWLEDGEMENT

A successful project is a fruitful culmination of efforts by many people, some directly involved and some others indirectly, by providing support and encouragement. Firstly I would like to thank the almighty for giving me the wisdom and grace for making my project a memorable one. I thank him for steering me to the shore of fulfillment under his protective wings.

I express my sincere gratitude to **Dr. T A Shahul Hameed**, Principal of T.K.M College of Engineering for giving me an opportunity to present my project. I would like to thank **Dr. Dimple A Shajahan**, Associate Professor and Head of the Department, CSE, TKMCE, for her constant support and encouragement throughout the work.

With a profound sense of gratitude, I would like to express my heartfelt thanks to my guide **Dr. Aneesh G Nath**, Associate Professor, CSE, TKMCE, and project coordinator **Dr. Ansamma John**, Professor, CSE, TKMCE for their expert guidance, cooperation and immense encouragement. I also extend my thanks to the entire faculty members and staffs of the Department of Computer Science & Engineering, TKMCE, who has encouraged me throughout this work.

I also express my thanks to my loving parents and friends, for their support and encouragement in the successful completion of this project work.

ASHA SIMON

**Abstract**

Today whatever look can see presence of digital image.Every day may groups like doctors, engineers, and students etc. release many images for their different needs.These images may contain both textual and non textual data. Text present in such images can provide meaningful information for content based repossession and many applications of computer vision like image under standing, reading number plates of moving vehicles and context retrieval for the investigation purposes.Text information found in these images may very different in fond size, style, alignment and orientation. In this way, it is very difficult to identify items with different characteristics and even more difficult to retrieve text if image is blurred or low resolution.There are many existing techniques for recover textual information from clear image. So in this paper, introduces a method to recover text from a blurred or low resolution image. The proposed technology is comprised of three main steps.(a) The deblurring process is applied to recover the clear image. (b) Extract text from images using text localization, segmentation and binarization techniques (c) Evaluation of proposed model. Input image debluring achieved by using super resolution convolution network. Text extraction can be achieved by using text extraction network. Here, use text detection to identify the text region on input image after that can find the exact position of text by using text localization and text segmentation separates the text from its background. Extensive experiments have been conducted on a synthetically generated dataset. Experimental results and analysis show that this system has better performance in terms of quantitative evaluation.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

when we capture an image text is common in natural scenes for example, sign boards, house number, posters and number plates of vehicle characters and Science in natural scene images provide important information for wide spectrum of applications such as seen understanding, context retrieval for many investigation purposes and for many other applications of computer vision. When capture an image it may blurred due to camera shake or motion of objects. In most cases we can not retake same image again. In these situations character recognition and image understanding from blurred image is more difficult. Not only the blur on image, text data on images also may varies because of difference in font size, style, alignment and orientation it makes the text detection and classification more challenging. Currently there are many methods for recover text data from a clear image . So in this paper introduces a technique to recover text from a blurred or low resolution image.

To extract the text from a low quality image, have to enhance the quality of the image to get the good result. However, in recent years techniques based on deep learning have been introduced which acquire classified features from training data. To obtain the characteristics of textual region, such as shape and edge of text, handcrafted features were also introduced. But instead of using low-quality photographs, the majority of them work with images of average quality.As mentioned above, text data provide important information. It is used in applications that use videos and images based on content, such as web image search based on content, information retrieval from videos, and text recognition and analysis based on mobile. Recently, many projects have used character localising method based on maximum stable extremal regions (MSER). Although MSER is considered as the best method for localizing text.The ability to read text data from low resolution scene text images helped in multiple real-world applications. For example, advantageous technologies for visually disabled persons and also in geo-localization, robot navigation and in autonomous vehicles as well as other edge and IoT based

units.So, most recently, computer vision has been paying more and more attention to the challenge of textual data recognition from low-quality natural scene text images. Scene text detection and classification became more difficult mainly because of two factors, known as internal factor and external factor. External factors means problems based on the environment, which causes blur, noise, occlusions etc. Internal factors means the dissimilarities in textual data from scene text images such as font size, colour variation, texture variation etc.Alternatively, most traditional methods such as MSC are very reliable in text detection and recently CNN-based methods have achieved very good results in text detection. Extracting features from the input we provide is an immense power of CNN-based models and they also help in learning high-level models. a novel methodology is suggested in this paper for the detection and classification of textual data from low quality natural scene images. The proposed technology is comprised of three main steps.(a) The deblurring process is applied to recover the clear image. (b) Extract text from images using text localization, segmentation and binarization techniques (c) Evaluation of proposed model. Input image debluring achieved by using super resolution convolution network. Text extraction can be achieved by using text extraction network. Here, use text detection to identify the text region on input image after that can find the exact position of text by using text localization and text segmentation separates the text from its background. Extensive experiments have been conducted on a synthetically generated dataset. Experimental results and analysis show that this system has better performance in terms of quantitative evaluation.

The remainder of this report is organized as follows. Chapter 2 recalls some related works used as reference for completing this study. Chapter 3 includes the proposed text data extraction from blurred image model is described. Chapter 4 presents the dataset used for conducting the experiment and results of the study. Finally, conclusion and future works are provided in chapter 5.

# Chapter 2

# Related Works

Hyukzae Lee et al. [1]suggested text-specific hybrid dictionary-based blind text picture deblurring technique. After thorough analysis, the text-specific hybrid dictionary has a high ability to offer solid contextual information for text picture deblurring. It is based on the observation that an intermediate latent picture includes both crisp and numerous different types of little blurred patches. gathered three distinct picture patch pairs to create the text-specific hybrid dictionary: Sharpsharp, motion blur-sharp, and sharp-sharp are the first three options. In this research, they offer a unique text picture deblurring technique employing a Text-specific Hybrid Dictionary to address the drawbacks of sparse representationbased. To deal with vast and complicated blurs, they use an iterative deblurring framework, which is a crucial component of unilateral dictionary learning techniques. They also create a dictionary utilising pairs of picture patches in a manner similar to linked dictionary learning ways to address the small and simple blurring in an intermediate latent image, which continue to be a significant issue in the unilateral dictionary learning approaches. overcome the problems of sparse representation-based approaches, in this paper,they propose a novel text image deblurring method using a Text-specific Hybrid Dictionary (THD). It is important to note that the proposed method fully takes the advantages of both coupled and unilateral dictionary learnings. They adopt an iterative deblurring framework, which is an essential workflow of the unilateral dictionary learning approaches, to cope with large and complex blurs. To deal with small and-simple blurs in an intermediate latent image, which still remain a serious problem in the unilateral dictionary learning approaches, Also construct a dictionary by using pairs of image patches similar to the coupled dictionary learning approaches In the new dictionary modeling mechanism, to obtain more text-like deblurring results, they also add image patch pairs of sharp textsharp text to reconstruct already sharpened patches. Here, it should be noted that each atom of hybrid dictionary contains a projection

matrix such that patches of intermediate latent images can be projected into sharp patches, which guarantees the desired properties of text image. A new optimization framework suitable for the task of text image deblurring is also proposed. They divide the problem into two subproblems: one is to estimate the latent image , and the other is to estimate the blur kernel.The iteratively optimize each of them by fixing the other.

Jian Sun et.al.[2] suggested a deep learning method employing a convolutional neural network to predict the probabilistic distribution of motion blur at the patch level (CNN). Here, we address the issue of predicting and removing non-uniform motion blur from a single fuzzy image. CNN made a prediction regarding the potential set of motion kernels. Then, to ensure smooth motion, a dense non-uniform motion blur field is inferred using a Markov random field model. Finally, motion blur is eliminated using a non-uniform deblurring model using patch-level image prior. In this work, they propose a novel deep learning-based approach to estimating non-uniform motion blur, followed by a patch statistics-based deblurring model adapted to non uniform motion blur. They estimate the probabilities of motion kernels at the patch level using a convolutional neural network, then fuse the patch-based estimations into a dense field of motion kernels using a Markov random field model.

To fully utilize the CNN, they propose to extend the candidate motion kernel set predicted by CNN using an image rotation technique, which significantly boost its performance for motion kernel estimation. Taking advantage of the strong feature learning power of CNNs, we can well predict the challenging non uniform motion blur that can hardly be well estimated by the state-of-the-art approaches They propose to estimate spatially-varying motion blur kernels using a convolutional neural network. The basic idea is that they first predict the probabilities of different motion kernels for each image patch. Then they estimate dense motion blur kernels for the whole image using a Markov random field model enforcing motion smoothness. They present an approach to predict motion blur kernels at the patch level. They decompose the image into overlapping patches of size 30$\times$ 30. A blurry patch is centered at pixel, they aim to predict the probabilistic distribution of motion kernels Taking the problem of motion kernel estimation as a learning problem, they utilize convolutional neural network to learn the effective features for predicting motion distributions. CNN model can predict the probabilities

of 73 candidate motion kernels. Obviously, they are not sufficiently dense in the motion space. They next extend the motion kernel set predicted by the CNN to enable the prediction for motion kernels outside.

A multiscale loss function mimics conventional coarse-to-fine methods. Also included is a brand-new, sizable dataset that features pairs of realistically fuzzy pictures and the ground-truth, sharp pictures that go with them. In place of any constrained blur kernel model, they suggest a multiscale CNN that directly restores latent pictures. This method does not estimate explicit blur kernels, in contrast to other methods. The artefacts that result from kernel estimation mistakes are not present in this method. Second, they significantly improve convergence by training the suggested model with a multi-scale loss suitable for a coarse-to-fine architecture. Additionally, they use adversarial loss to enhance the results even more. Thirdly, they suggest a brand-new dataset of realistically fuzzy photos with sharp ground-truth images.

Therefore, all the existing methods still have many problems before they could be generalized and used in practice. These are mainly because of the use of simple and unrealistic blur kernel models. Thus, to solve those problems, in this work, they discuss about a novel end-to-end deep learning approach for dynamic scene deblurring. First, they propose a multi-scale CNN that directly restores latent images without assuming any restricted blur kernel model. Especially, the multi-scale architecture is made to seem like conventional coarse-to-fine optimization techniques. Unlike other approaches, this method does not estimate explicit blur kernels. Accordingly, this method is free from problems that happened due to kernel estimation errors. Second, they train the proposed model with a multi-scale loss that is suitable for coarse-to-fine architecture that considerably enhances convergence. In addition, they further improve the results by employing adversarial loss. Third, they suggest a new real world blurry image dataset with ground truth sharp images. To obtain kernel model free dataset for training, they employ the dataset acquisition method Since the integration of sharp images during the shutter period can be used to model the blurring process. They used a high-speed camera to record a series of sharp images of a dynamic scene, and averaged them to generate a blurry image by considering gamma correction.

Jinshan Pan Et.al [4] presented a straightforward but efficient L0-regularized prior for text picture deblurring based on intensity and gradient. For pro-

ducing reliable kernel estimate intermediate results, a powerful optimization technique is applied. The suggested method does not use any sophisticated filtering techniques, which are essential in modern deblurring algorithms, to locate salient edges. Here, a quick way to eliminate artefacts and better superior deblurred images is obtained for the last step of latent image restoration. The proposed L0 intensity and gradient prior is based on the observation that, text region and background of blurred image usually have similar uniform intensity values than clean images. If you only take zero peaks into account, the pixel values of text images are quite sparse. The histogram of pixel intensity for a blurred text image differs from that of a clear image. Most significantly, it lacks the zero peak. Specifically, images containing blurred text have denser pixel values. In this formulation, they use this intensity attribute, which is general for text images, as one regulation factor. It is well known that estimating motion kernels from blurred images with saturated pixel regions is a complex issue.

Although some non-blind deblurring methods have been proposed, it remains challenging to develop effective blind deblurring algorithms. Saturated regions usually appear rarely in clear images and these regions are much larger in the low resolution images. As the L0 norm used in the newly proposed algorithm is same to an adaptive hard threshold strategy. If look at the binary images of clear and corresponding blurred images, the binary image of blurred image has more nonzero elements than the binary of clear image. The suggested deblurring algorithm favours solutions with fewer blobs or streaks in the latent clear pictures since the L0 norm reduces the number of nonzero coefficients. In this essay, they suggest a straightforward but powerful prior for text image deblurring. Although the suggested prior is using the characteristics of two-tone text graphics, it can also be used successfully for images with large paragraphs and low-illumination scenes with saturated areas. They provide an efficient optimization technique based on a half quadratic splitting strategy model using this prior, which guarantees that each sub-problem has a closed form solution. The suggested approach does not need any complex processing techniques like filtering and adaptive segmentation.

Jose Jaena Mari Ople et al.[5] Multi-scale characteristics have been suggested to offer the spatial dependencies required to deblur irregular blurs. Here, two techniques are employed to extract multi-scale features efficiently:

a coarse-to-fine method that applies the network to various picture sizes to extract multi-scale features, and dilated convolutions that employ varying dilation rates to extract multi-scale features. Combining the two methods has a multiplicative effect since multi-scale features from dilated convolutions are produced from the input images at multiple scales. In order to speed up the deblurring process, parallel convolutions are also included in the network design. They employ the Scale Recurrent Network design, which is based on a U-Net architecture and the coarse-to-fine scheme, to depict the difficult mapping from a hazy image to a sharp image with fewer parameters. Due to downsampling, spatial information is lost in the encoder block, and if more multi-scale features are required, adding more scale levels in the coarse-to-fine scheme is inefficient. The architecture of these deep learning approaches typically has many convolutional layers because of the following factors.

Wenqi Ren Et. al [6] presented a neural network with spatial variation to deblurred dynamic scenes. The recurrent neural network is used to perform deconvolution operator on feature maps that derived from the input picture. Another CNN is used to learn the spatially variable weights of the RNN. The RNN can implicitly model the deblurring process using spatially varying kernels as a result of its spatial awareness. For deblurring, one-dimensional and two-dimensional RNNs are constructed to better utilise the capabilities of the spatially variable RNN. The third component, which is based on a CNN, reconstructs the final deblurred feature maps into a restored image. Additionally, the entire network can undergo end-to-end training. To better exploit properties of the spatially varying RNN, they exploit both one-dimensional (1D) and two-dimensional (2D) RNN for deblurring. Every pixel in 1D RNNs is connected to just one pixel from the previous row or column, and thus the relationship between neighbouring pixels are underutilized. While 2D RNNs are better than 1D RNNs at learning the denser propagation between adjacent pixels and capturing a larger receptive region.

Wenqi Ren Et. al [6] presented a neural network with spatial variation to deblurred dynamic scenes. The recurrent neural network is used to perform deconvolution operator on feature maps that derived from the input picture. Another CNN is used to learn the spatially variable weights of the RNN. The RNN can implicitly model the deblurring process using spatially varying kernels as a result of its spatial awareness. For deblurring, one-dimensional and two-dimensional RNNs are constructed to better utilise the capabilities of

the spatially variable RNN. The third component, which is based on a CNN, reconstructs the final deblurred feature maps into a restored image. Additionally, the entire network can undergo end-to-end training. To better exploit properties of the spatially varying RNN, they exploit both one-dimensional (1D) and two-dimensional (2D) RNN for deblurring. Every pixel in 1D RNNs is connected to just one pixel from the previous row or column, and thus the relationship between neighbouring pixels are underutilized. While 2D RNNs are better than 1D RNNs at learning the denser propagation between adjacent pixels and capturing a larger receptive region.

Xirui Yang Et.al[7] suggested a convolutional neural network-based end-to-end picture blind deblurring technique. This method uses three separate convolutional neural networks: a deblurring network, a super-resolution network, and a feature fusion network to combine picture deblurring with super-resolution reconstruction. In order to improve network performance, it is built on Res2Net, densely coupled convolutional networks, and the segmentation channel method. The Dense-Net and Res2Net are utilised as the fundamental modules that make up the algorithm neural network, and the network is converted to an encoding decoding structure so that the feature information between the encoding layer and the decoding layer may be exchanged. This algorithm produced promising experimental outcomes. The experimental findings demonstrate that this technique likewise struggles to effectively handle complicated non-uniform fuzzy objects. The validity of the feature fusion module is checked concurrently.

The EDSR is currently the best method for indexing in the field of natural image super-resolution. For the sake of fairness, when retraining the neural network of the EDSR, the same learning rate, iteration number, batch size and other hyper parameters as those proposed in this section are set. The experimental results show that this method also can't deal with complex non-uniform fuzzy artifacts well. At the same time, the validity of the feature fusion module is verified. Its resolution is improved compared with the previous two figures, but it is still relatively smooth at the local details. The boundary distinction is not obvious.In summary, the proposed algorithm has a significant effect on the improvement of visual quality of ultrasound endoscopic images.

A.Chakrabarti [8] Describe a novel approach to blind motion deblurring that makes use of a neural network that has been trained to predict the

sharpness of picture patches from observations blurred by an unknown motion kernel. This network learns to forecast the complex Fourier coefficients of a deconvolution filter that will be applied to the input patch for restoration rather than directly regressing to patch intensities. In order to create an initial approximation of the crisp image, they independently apply the network to all overlapping patches in the observed image. They then execute non-blind deconvolution using this kernel after explicitly estimating a single global blur kernel and linking it to the observed image. The accuracy and robustness of this method are comparable to those of modern iterative methods, whereas being paralleled on GPU hardware significantly faster.

In this paper,they propose a new approach for blind deconvolution of natural images degraded by arbitrary motion blur kernels due to camera shake. At the core of our algorithm is a neural network trained to restore individual image patches. This network differs from previous architectures in two significant ways: Rather than formulate the prediction task as blur kernel estimation through iterative refinement, or as direct regression to deblurred intensity values, network is trained to output the complex Fourier coefficients of a deconvolution filter to be applied to the input patch.Multi-resolution frequency decomposition is used to encode the input patch, and limit the connectivity of initial network layers based on locality in frequency . This leads to a significant reduction in the number of weights to be learned during training, which proves crucial since it allows us to successfully train a network that operates on large patches, and therefore can reason about large blur kernels For whole image restoration, the network is independently applied to every overlapping patch in the input image, and its outputs are composed to form an initial estimate of the latent sharp image. Despite reasoning with patches independently and not sharing information about a common global motion kernel, found that this procedure by itself performs surprisingly well.

M.Hradis et.al[9]deal with the blind deconvolution and denoising issue. They concentrate on text document restoration and demonstrate that a convolutional neural network can successfully restore this kind of highly structured material. Without assuming any particular blur and noise models, the networks are trained to reconstruct high-quality images directly from hazy inputs. They use a sizable collection of written documents and a combination of realistic defocus and camera shake blur kernels to show how well the convolutional networks function. On this synthetic data, convolutional networks

perform noticeably better in terms of image quality and OCR accuracy than current blind deconvolution techniques, even ones that are tuned for text. In fact, for all noise levels save the lowest, the networks perform better than even the most advanced non-blind techniques. The method is tested using actual images captured by various devices.

They directly predict clean and sharp images from corrupted observed images by a convolutional network.The architecture of the networks is inspired by the recently very uccessful networks that redefined state-of-the-art in many computer vision tasks includingobject and scene classification , object detection, and facial recognition.All these networks are derived from the ImageNet classification network They tested the approach on the task of blind deconvolution with realistic defocus and camera shake blur kernels on a large set of documents from the CiteSeerX1repository. They explored different network architecture choices, and compared results to state-of-the-art blind and non-blind deconvolution methods in terms of image quality and OCR accuracy. Purposely limited the image degradations to shift-invariant blur and additive noise to allow for fair comparison with the baseline methods, which are not designed to handle other aspects of the image acquisition process. To validate our approach, we qualitatively evaluated the created networks on real photos of printed documents.

# Chapter 3

# Methodology

A process plan of the proposed framework is presented in Fig. 3.1 The

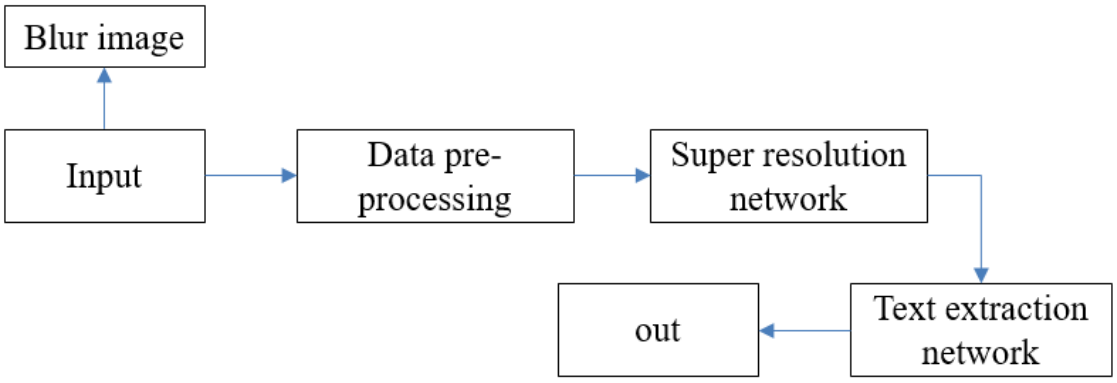


Figure 3.1: framework to extract text from blurred image.

proposed work consists of three parts: (1) Following the preprocessing of low-quality images,(2) Super resolution network (3) text extraction network. First preprocessing the input data then preprocessed data passed to super-resolution network output from the super-resolution network is the set of the deblurred images. The given set of output then passed to the text extraction network.The final output from the network is set of editable text.

## 3.1 Data Collection

Here needs the set of clear image and corresponding blurred image as dataset to train the system.Here dataset is created by collecting the set of images of pdf notes, sign boards and numberplates of vehicles using the camera and then applying Gaussian blur to those images. The proposed methodologyuses the dataset created in this way. It contains 1010 pairs of blur and sharp images. Out of which 700 pairs for training and 300 pairs for testing. A sample of dataset is shown in figure 3.2.

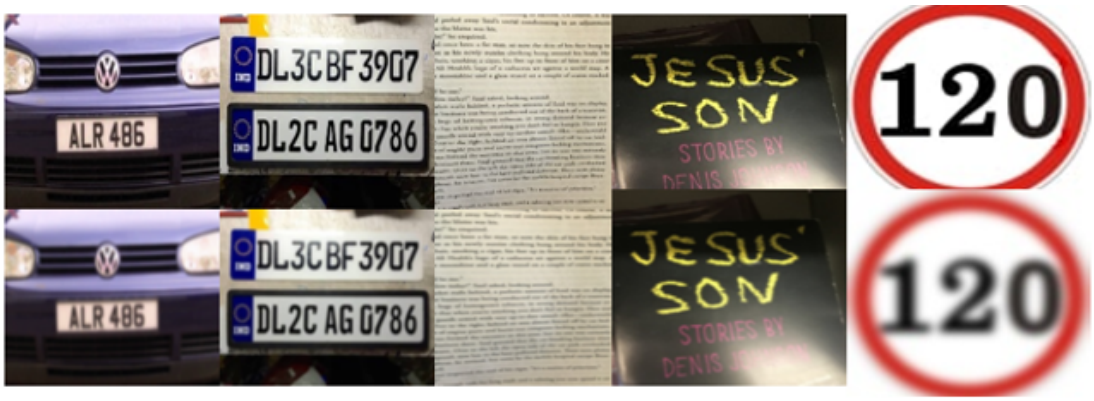


Figure 3.2: A sample of dataset.

## 3.2 Input Data Preprocessing

The collected data is given for data preprocessing. Image preprocessing are the steps taken to format images before the used. It includes image resizing,reshaping and normalization. In this images are in RGB colour space with resolution 260*190. The major objectives of utilising any sort of neural network are better prediction outcomes and improved accuracy, and improving input data quality for neural networks can significantly contribute to these outcomes. The preprocessing procedures comprise: putting each photograph in its original format. picture cropping to remove unused areas. converting them to numbers so that computers can use them to learn (array of numbers). Depending on the image resolution, computers interpret an input image as a collection of pixels.

## 3.3 Splitting of Dataset into Test and Training Data

Splitting the data into training and testing helps to validate the model. Training data is used to train or fit the model which is to be used.It helps the model to learn how to make predictions for a given task. Once the model is trained with training dataset then the model has to be tested using test dataset. Here it contains 1000 pairs of blur and sharp images. Out of which 700 pairs for training and 300 pairs for testing.

```
[[[[-0.41960784 -0.39607843 -0.44313725]
   [-0.67058824 -0.64705882 -0.75686275]
   [-0.61568627 -0.59215686 -0.75686275]
   ...
   [-0.45098039 -0.34901961 -0.30196078]
   [-0.62352941 -0.41960784 -0.30196078]
   [-0.58431373 -0.2627451  -0.12156863]]

  [[-0.63921569 -0.61568627 -0.69411765]
   [-0.65490196 -0.63921569 -0.74117647]
   [-0.59215686 -0.57647059 -0.73333333]
   ...
   [-0.42745098 -0.34117647 -0.30196078]
   [-0.48235294 -0.29411765 -0.2       ]
   [-0.5372549  -0.24705882 -0.12156863]]
```

Figure 3.3: Preprocessed data of image.

## 3.4 Super Resolution Network

The preprocessed image is then fed to the image super resolution network to enhance the quality of input blurred image. super resolution CNN is a feed-forward artificial neural network with a Wavelet filter.

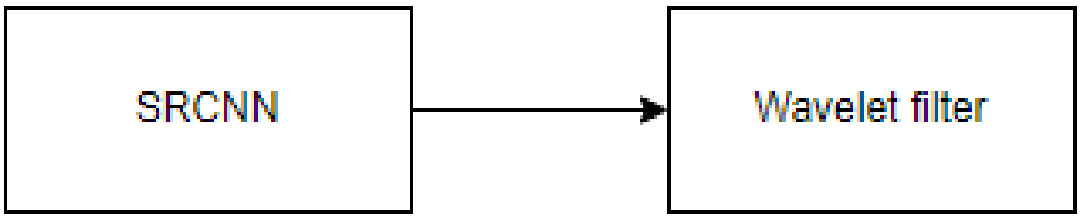


Figure 3.4: Super Resolution Network architechure.

### 3.4.1 Super resolution convolutional Neural Network

Super resolution convolutional Neural Network (SRCNNs) consist of multiple layers of small neuron collections that process the receptive fields ie, portions of the input image, when used for image recognition. To obtain a better rep-

resentation of the original image, outputs of these collections are then tiled so that they overlap and this is repeated for every layer.Major advantages of CNNs are the lack of dependence on prior knowledge and human effort in designing features. CNN architecture is formed by arranging a set of distinct layers that transform the input into an output through a differentiable function.

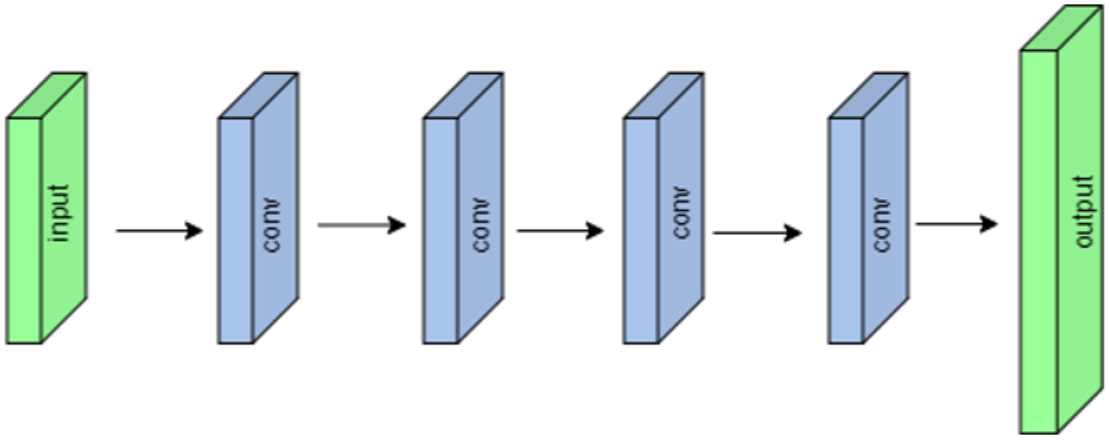


Figure 3.5: Image super Resolution Network.

The CNN which is used for image deblurring in this technique is shown in Figure 3.4. In this structure of CNN, six convolutional layers including input layer along with ReLU layers.Usually while designing a CNN pooling layer is used but, here pooling layer is not used so get the output image with resolution same as input image.In this network end to end mapping between high resolution and low resolution image is takes place. The main building block of a CNN is the convolutional layer. All learning process achieved via hidden layers, feature extraction and aggregation are also formulated as convolutional layer. For Non linear mapping uses RELU as activation function this operation map high dimensional vector inti another high dimensional vector then each mapped vector is the representation of high resolution patch. The main advantage of using the ReLU function over other activation functions is that it does not activate all the neurons at the same time.

Learning the end-to-end mapping function $F$ requires estimation of weight and bias, this is achieved through minimising loss between reconstructed image F(Y,$\Theta$) and corresponding ground truth high resolution image (X). In this used mean squared error as loss function, corresponding is shows in

equation 3.1.

$$L(\Theta) = \frac{1}{n}\sum_{i=1}^{n} \|F(\mathbf{Y}_i;\Theta) - \mathbf{X}_i\|^2 \tag{3.1}$$

Here $n$ represents the number of training samples, F(Y,Θ) represents reconstructed image and $X$ represents the ground truth high resolution image. Losses are calculated and then minimized using standard propagation. Here introduces a model which is very simple in structures with a moderate number of filters and layers, so it has lot of advantages because of this simple in structure one is, this achieves faster speed for online usage even on CPU. Second one, is this is fully feed forward and does not need to solve any Optimisation problem last one is, gives the better restoration quality.

Output from super resolution CNN (SRCNN) is set of deblurred images, sample is shown in figure 3.6

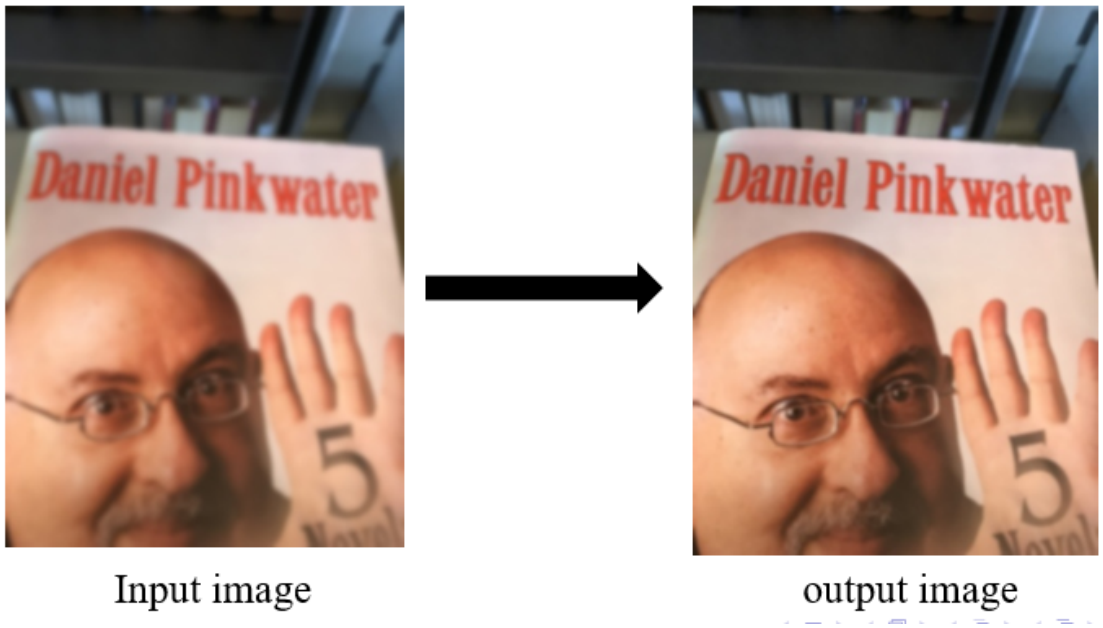


Figure 3.6: Intermediate output from super resolution network.

### 3.4.2 Wavelet Filter

wavelet de-noising is based on the wavelet representation of the image. First Gaussian noise is represented by small values of wavelet domain and that

removed by setting coefficients below given threshold to zero.Commonly two type of threshold can used to setting coefficients below given threshold.Fist one is hard threshold in this case threshold is shrinking towards the zero. Second one is soft threshold. Here uses soft threshold method. Figure 3.7 shows output from super resolution network.Its quality is more good than intermediate out from SRCNN. To compare intermediate out from SRCNN

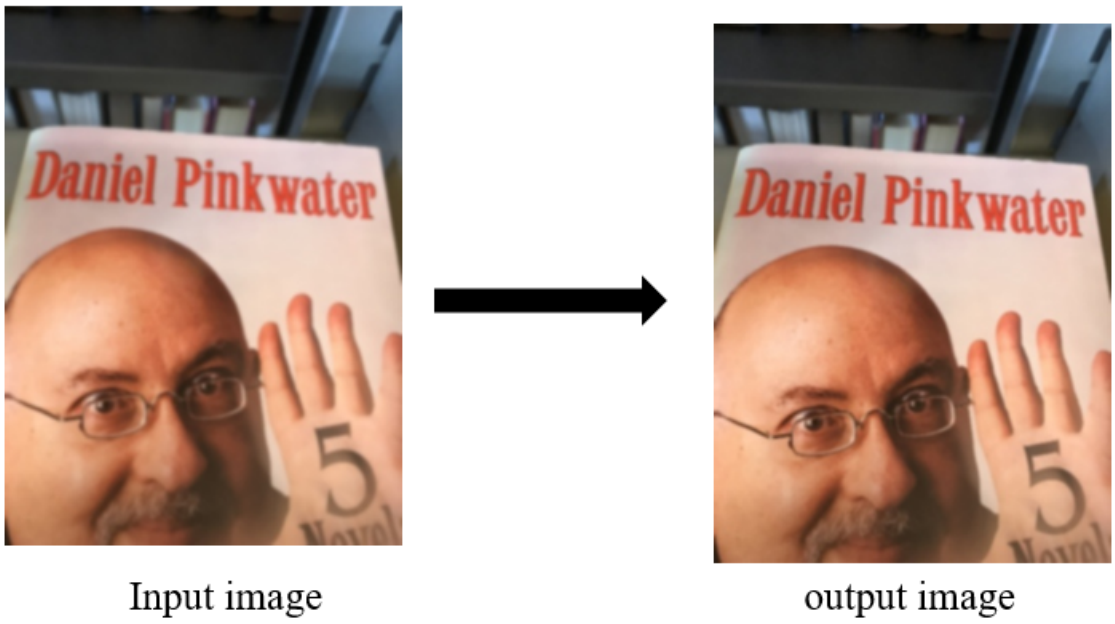


Figure 3.7: output from super resolution network.

to final output here uses BRISQUE value and compared out is shown in figure 3.8. BRISQUE value is usually in the range of [0-100].Zero means better quality and hundred represents lower quality.According to BRISQUE value analysis lower value reflects better quality of image with respect to the input image. From the figure 3.8 can understand that the out put from the super resolution network is far better than the intermediate output. Super

```
BRISQUE value for output :  29.403604768162353
BRISQUE value for input :  41.2374124413071
```

Figure 3.8: output from super resolution network.

resolution network is only a part of overall blurred text extraction network.

The importance of compare the intermediate output and final output is to shows the importance of use of wavelet filter with SRCNN.

## 3.5 Text Extraction Network

Third part of the architecture of text extraction from blurred image is text extraction network shown in figure 3.9. This network is used to extract the text from output of super resolution network. Output of super resolution network act as the input of text extraction network.

Various steps involved in text extraction network are

- Converting coloured image to gray scale image.
- Segmentation.
- Pass the segmented image to trained CNN.

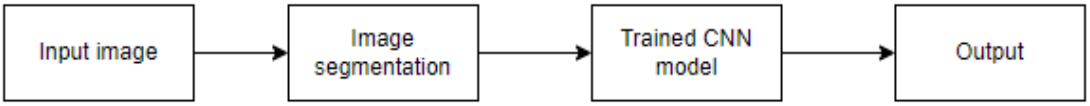


Figure 3.9: Architecture of text extraction network.

### 3.5.1 Converting Coloured Image to Gray Scale Image.

A digital color image is a colour image that includes color information for each pixel. There are various color models which are used to represent a color image. These are RGB color model, in which red , green and blue light is added together in various ways to reproduce a broad array of colors. The other models are CMY color model which uses cyan, magenta and yellow light and HSI model which uses hue , saturation and intensity variations. Gray scale images have range of shades of gray without apparent color. These are used as less information needs to be provided for each pixel. In an 8 bit image, 256 shades are possible. The darkest possible shade black is represented as 00000000 and lightest possible shade white is represented as 11111111. One of the advantage of convert colour image to grey scale is dimension reduction. For example, In RGB image there are three colour channels and

three dimensions while grey scale images are single dimension and it helps to reduces the model complexity.

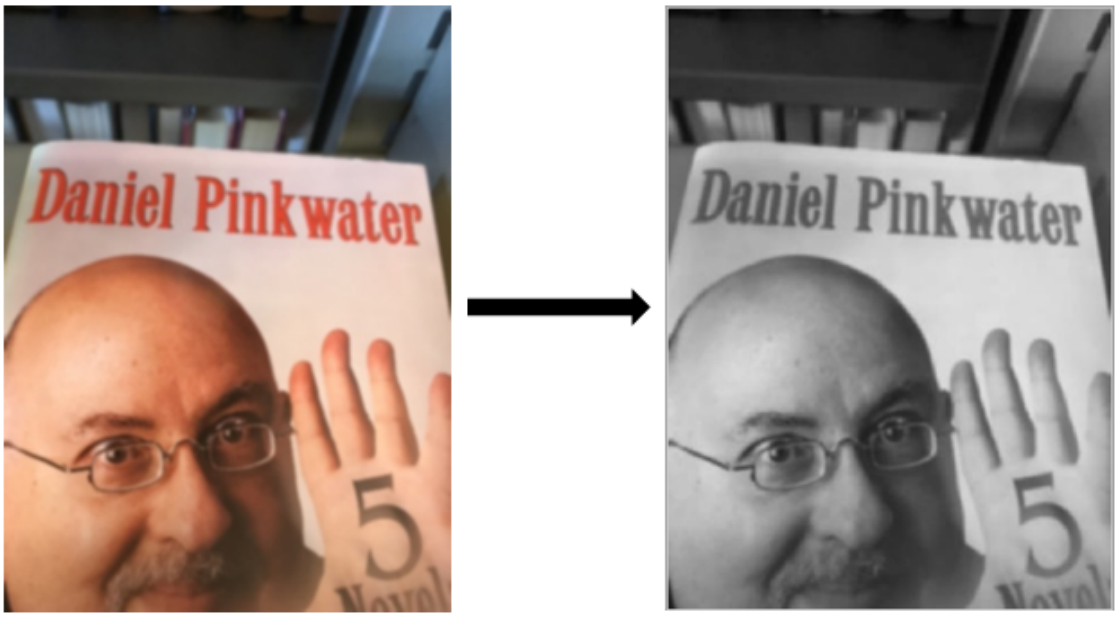


Figure 3.10: Grey scale of input image.

### 3.5.2 Segmentation

The converted grey scale image is next passed to segmentation process. Binarization method that is used for segmentation process. A Binary image is a digital image that can have only two possible values for each pixel. Each pixel is stored as single bit 0 or 1. The name black and white is often used for this concept.To form a binary image first select a threshold intensity value. All the pixels having intensity greater than the threshold value are changes to 0 that is black and the pixels with intensity value less than the threshold intensity value are changed to 1 that is white. Thus the image is changed to a binary image. Goal of image binarization is the segmentation of document into foreground text and background. For this here uses morphological operation.

### 3.5.3 Input to Character Recognition Network

The binary image get in above is passed to the character recognition network to identify the highlighted text. Here building character recognition network

using convolutional neural network.

Pattern recognition is extremely difficult to automate. Human recognize various objects and identify that from the large amount of visual information, it requiring very little effort. Simulating the task performed by human to identify to the extent allowed by physical limitations will be very dificult for the system. This necessitates study and simulation of Artificial Neural Network. In Neural Network, each node perform some simple computation and each connection conveys a signal from one node to another labeled by a number called the weight indicating the extent to Different choices for weight results in different functions are being evaluated by the network. If in a given network whose weight are initial random and given that all know the task to be accomplished by the network , a learning algorithm must be used to determine the values of the weight that will achieve the desired task. Learning Algorithm qualifies the computing system to be called Artificial Neural Network. The node function was predetermined to apply specific function on inputs imposing a fundamental limitation on the capabilities of the network. Typical pattern recognition systems are designed using two pass. The first pass is a feature extractor that finds features within the data which are specific to the task being solved The second pass is the classifier, which is more general purpose and can be trained using a neural network and sample data sets. Clearly, the feature extractor typically requires the most design effort, since it usually must be hand crafted based on what the application is trying to achieve.

This module require identification of characters. The behaviour of dataset is changes according to the application. Here uses the set of image of alphabets to train the system. In first phase preprocessing the given input text image for separating the Characters from it and normalizing each characters. Initially specify an input image file, which is opened for reading and preprocessing.

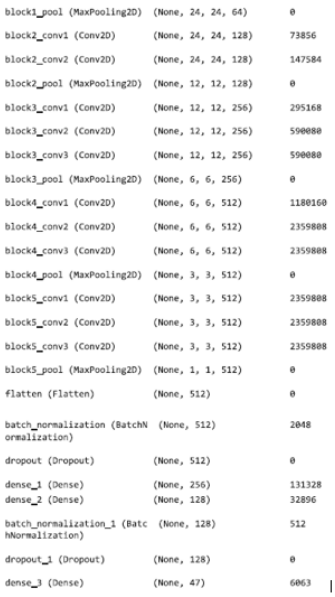


Figure 3.11: Text extraction network structure.

In this all given images are in gryscale, so no need of convertion into binary format. Now extracted the character after that need to normalize the size of the extracted characters. There are large variations in the sizes of each extracted character and input hence need a method to normalize the size. Figure 3.11 shows the layer structure of text extraction network. Output from this module is set of text exxtracted fro input text image.

# Chapter 4

# Experimental Results and Discussions

The text is extracted using Text extraction from blurred image architecture which uses synthetic dataset.For an input image the output obtained is shown in figure 4.1. The loss graph is shown in figure 4.2. HereMean squared


Figure 4.1: Extracted text.

error (MSE) loss functions is used. Objective of loss function is measure how good the prediction model does in terms of being able to predict the expected outcome.

From the results, the following observations are made:

- A small number of nodes in the hidden layer lower the accuracy.
- A large number of neurons in the hidden layer help to increasing the accuracy.

| Epochs | Accuracy |
|---|---|
| 300 | 31 |
| 600 | 65 |
| 1000 | 89 |

Table 4.1: Accuracy

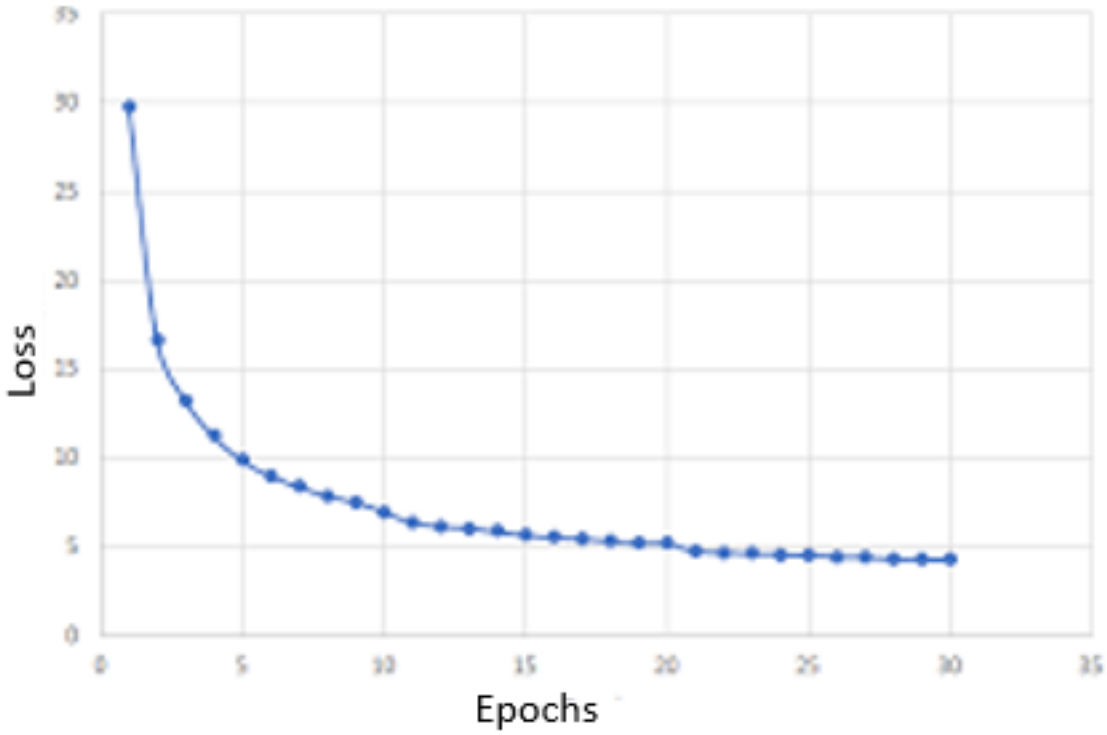


Figure 4.2: Loss curve.

- Increase in number of cycles increase accuracy.
- Accuracy also can increased by increasing the size of the training set.

# Chapter 5

# Conclusion and Future Work

In this paper, Deep learning technique to extract text from low quality natural scene images is presented.The query image is synthetically blurred using Gaussian filter. Further deblurring methos using super resolution CNN adapted for enhancing contrast quality of deblurred of image. Then Text extraction is applied for localizing and detecting text regions, and non-text areas are discarded. Morphological operation is applied for better separation of foreground from background and it also has less false positive rate. In this process can not apply text extraction properly on images which contain large paragraphs. This problem can also be solved in future work by using SWT technique on images to improve the results. The classification results are obtained on various distributions of training and testing sets. Furthermore, a different scenario that is separately extracted features and CNN features is also adapted to monitor the credibility of the CNN model. It is concluded that the proposed methodology works fine and responds well to all types of tests. Finally, it is observed that the proposed work performs well in detecting text.

# References


[1] .Chanho Jung, and Changick Kim."Blind Deblurring of Text Images Using a Text-Specific Hybrid Dictionary". IEEE Transaction Image Processing, VOL. 29, 2020.

[2] .S. Nah, T. H. Kim, and K. M. Lee. "Deep multi-scale convolutional neural network for dynamic scene deblurring". IEEE Conference on Computer Vision and Pattern Recognition (CVPR) pages 3883–3891, 2017

[3] J. Sun, W. Cao, Z. Xu, and J. Ponce. "Learning a convolutional neural network for non-uniform motion blur removal". In CVPR, pages 769–777. IEEE, 2015.

[4] J. Pan, Z. Hu, Z. Su, and M.-H. Yang,."Deblurring text images via L0-regularized intensity and gradient prior" in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 2901–2908

[5] Jose Jaena Mari Ople1, Pin-Yi Yehi." Multi-Scale Neural Network with Dilated Convolutions for Image Deblurring". IEEE Access, 8, 53942-53952 16th March 2020.

[6] Wenqi Ren,et al."Deblurring Dynamic Scenes via Spatially Varying Recurrent Neural Networks". DOI10.1109/TPAMI.2021.3061604, IEEE Transactions on Pattern Analysis and Machine Intelligence

[7] Xirui Yang,et al."Endoscopic Image Deblurring and Super-Resolution Reconstruction Based on Deep Learning". International Conference on Artificial Intelligence and Computer Engineering (ICAICE) 2020

[8] A. Chakrabarti, "A neural approach to blind motion deblurring," in European Conference on Computer Vision, 2016.

[9] M. Hradis, J. Kotera, P. Zemcık, and F.Sroubek, "Convolutional neural networks for direct text deblurring," in British Machine Vision Conference, 2015.

[10] H. Arshad, M. A. Khan, M. I. Sharif, M. Yasmin, J. M. R. S. Tavares et al. (2020). , "A multilevel paradigm for deep convolutional neural network features selection with an application to human gait recognition," Expert Systems , vol. 21 , no.3 , pp. e12541.

[11] M. I. Sharif, J. P. Li, M. A. Khan and M. A. Saleem. (2020). "Active deep neural network features selection for segmentation and recognition of brain tumors using MRI images," Pattern Recognition Letters , vol. 129, pp. 181–189.

[12] M. Rashid, M. A. Khan, M. Alhaisoni, S. H. Wang and S. R. Naqvi. (2020). "A sustainable deep learning framework for object recognition using multi-layers deep features fusion and selection," Sustainability , vol. 12 , no. 12 , pp. 5037.

[13] D. Karatzas, L. Gomez-Bigorda, A. Nicolaou, S. Ghosh, A. Bagdanov et al. (2015). , "ICDAR 2015 competition on robust reading," in 2015 13th Int. Conf. on Document Analysis and Recognition, Tunis, Tunisia, pp. 1156–1160.

[14] D. Karatzas, F. Shafait, S. Uchida, M. Iwamura, L. G. i Bigorda et al. (2013). , "ICDAR 2013 robust reading competition," in 2013 12th Int. Conf. on Document Analysis and Recognition, Washington, DC, USA, pp. 1484–1493.

[15] A. Shahab, F. Shafait and A. Dengel. (2011). "ICDAR, 2011 robust reading competition challenge 2: Reading text in scene images," in 2011 Int. Conf. on Document Analysis and Recognition, Beijing, China, pp. 1491–1496.

[16] M. E. Maros, C. G. Cho, A. G. Junge, B. Kämpgen, V. Saase et al. (2020). , "Comparative analysis of machine learning algorithms for computer-assisted reporting based on fully automated cross-lingual RadLex® mappings,".

[17] Y. Liu, C. Yang, L. Jiang, S. Xie and Y. Zhang. (2019). "Intelligent edge computing for IoT-based energy management in smart cities," IEEE Network , vol. 33 , no. 2 , pp. 111–117.

[18] R. S. Alonso, I. Sittón-Candanedo, Ó García, J. Prieto and S. Rodríguez-González. (2020). "An intelligent edge-IoT platform for monitoring livestock and crops in a dairy farming scenario," Ad Hoc Networks , vol. 98 pp. 102047. 10 .

[19] P. Lyu, C. Yao, W. Wu, S. Yan and X. Bai. (2018). "Multi-oriented scene text detection via corner localization and region segmentation," in Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Beijing, China, pp. 7553–7563.

[20] M. A. Khan, K. Javed, S. A. Khan, T. Saba, U. Habib et al. (2020). , "Human action recognition using fusion of multiview and deep features: An application to video surveillance," Multimedia Tools and Applications , vol. 10 , pp. 335

[21] A. Majid, M. A. Khan, M. Yasmin, A. Rehman, A. Yousafzai et al. (2020). , "Classification of stomach infections: A paradigm of convolutional neural network along with classical features fusion and selection," Microscopy Research and Technique , vol. 83 , no. 5 , pp. 562–576. 13

[22] M. A. Khan, M. A. Khan, F. Ahmed, M. Mittal, L. M. Goyal et al. (2020). , "Gastrointestinal diseases segmentation and classification based on duo-deep architectures," Pattern Recognition Letters , vol. 131 , pp. 193–204. 14

[23] F. E. Batool, M. Attique, M. Sharif, K. Javed, M. Nazir et al. (2020). , "Offline signature verification system: A novel technique of fusion of GLCM and geometric features using SVM," Multimedia Tools and Applications , pp. 1–20.

[24] T. Akram, M. Sharif and T. Saba. (2020). "Fruits diseases classification: Exploiting a hierarchical framework for deep features fusion and selection," Multimedia Tools and Applications , pp. 1–21.

[25] A. Adeel, M. A. Khan, T. Akram, A. Sharif, M. Yasmin et al. (2020). , "Entropy-controlled deep features selection framework for grape leaf diseases recognition," Expert Systems , vol. 1 , no. 1 , pp. 1.

[26] Q. Yang, M. Cheng, W. Zhou, Y. Chen, M. Qiu et al. (2018). , "Inceptext: A new inception-text module with deformable psroi pooling

for multi-oriented scene text detection," arXiv preprint arXiv: 1805. 01167.

[27] M. S. Das, B. H. Bindhu and A. Govardhan. (2012). "Evaluation of text detection and localization methods in natural images," International Journal of Emerging Technology and Advanced Engineering , vol. 2 , no. 6 , pp. 277–282.

[28] A. Criminisi, P. Pérez and K. Toyama. (2004). "Region filling and object removal by exemplar-based image inpainting," IEEE Transactions on Image Processing , vol. 13 , no. 9 , pp. 1200–1212.

[29] J. Kostková, J. Flusser, M. Lébl and M. Pedone. (2019). "Image invariants to anisotropic Gaussian blur," in Scandinavian Conf. on Image Analysis, Cham: Springer, pp. 140–151.

[30] Seaboyer J, Barnett T. New perspectives on reading and writing across the disciplines[J]. Higher Education Research Development, 2019, 38(1):1-10.

[31] Montserrat D M, Lin Q, Allebach J, et al. Logo detection and recognition with synthetic images[J]. Electronic Imaging, 2018, 2018(10):3371-3377.

[32] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic data for text localisation in natural images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2315–2324.

[33] W. Liu, C. Chen, and K.-Y. K. Wong, "Char-Net: A character-aware neural network for distorted scene text recognition," in Proc. AAAI Conf. Artif. Intell., New Orleans, LA, USA, 2018, pp. 7154–7161.

[34] P. He, W. Huang, Y. Qiao, C. C. Loy, and X. Tang, "Reading scene text in deep convolutional sequences," in Proc. AAAI, vol. 16, 2016, pp. 3501–3508.

[35] V. K. Kurmi, S. Kumar, and V. P. Namboodiri, "Attending to discriminative certainty for domain adaptation," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 491–500.