AI based Document Image Processing for OCR(Robust OCR)

Ashutosh Singh N.Sai Satya Harika Piyush Sharma S Sarathchandran Gade Manasa

***Abstract*—Document image analysis refers to algorithms and techniques that are applied to images of documents to obtain a computer-readable description from pixeldata.Awell-knowndocumentimageanalysisproductistheOpticalCharacter Recognition (OCR) software that recognizes characters in a scanned document. OCR makes it possible for the user to edit or search the document's contents. In this paper we briefly describe various components of a document analysis system. Many of these basic building blocks are found in most document analysis systems, irrespective of the particular domain or language to which they are applied. We hope that this paper will help the reader by providing the background necessary to understand the detailed descriptions of specific techniques presented in other papers in this issue.**

***Index Terms*—OCR; feature analysis; document processing; graphics recognition; character recognition; layout analysis.**

I. Introduction

The field of pattern recognition has become one of the broad areas where more and more researchers have worked. The goal of researchers in this field is to find algorithms that can solve on a computer the problems of pattern recognition, which are intuitively resolved by humans. Optical character recognition (OCR) is one of the fields in pattern recognition; its purpose is to transform an image of handwritten, typewritten or printed text into an understandable representation that a computer can easily recognize. Consequently, the OCR system is applied in several applications in various domains such as: bank check processing, postal code recognition, mail sorting, digital libraries, security system, etc. Optical character recognition (OCR)systems face severe challenges in recognizing images that lack resolution. Most OCR models trained on high resolution (HR) text images fail in case of low resolution (LR), noisy or noisy-LR text images. Super-resolution (SR) processes LR text images and de-noising algorithm is applied on noisy text images for accurate text recognition. But pre-processing of the noisyLR text images is a challenging problem [1]. The traditional approach of pre-processing these images includes de-noising followed by super-resolution. But this approach results in loss of textural and high-frequency details during the de-noising process and the SR process is then not able to accurately super-resolve the image. Thus, the resultant HR image has some important details missing leading to low recognition accuracies. So we propose a frame work that super resolve the less noisy images. Since every OCR engine cannot be expected to be trained on all resolutions, thus a noise resilient super-resolution module as a part of the end-to-end OCR framework can make a generic architecture.

II. Related work

A. Factors Affecting the Text Recognition Quality

Many factors influence the precision of character recognized using OCR. The factors are scan resolution, scanned image quality, printed documents category either photocopied or laser printer, quality of the paper, and linguistic complexities. The uneven illumination and watermarks are few factors faced in OCR system that influence the accuracy of OCR.

B. Significance of Preprocessing in Text Recognition

The preprocessing step is necessary to obtain better text recognition rate, using efficient algorithms of preprocessing creates the text recognition method robust using noise removal, image enhancing process, image threshold process, skewing correction, page and text segmentation, text normalization and morphological operations.

C. Denoising images

Autoencoders have been widely applied in dimension reduction and image noise reduction.Modeling image data requires a special approach in the neural network world. The best known neural network for modeling image data is the Convolutional Neural Network (CNN, or ConvNet) or called Convolutional Autoencoder.It helps in denoising of images.

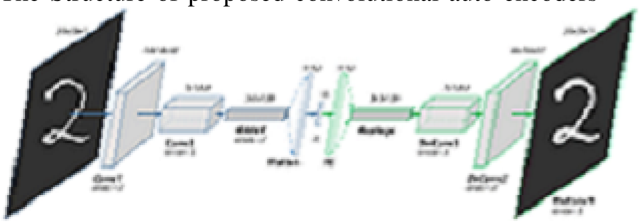
There have been works around text image super-resolution [5], [6], [7], [8] and denoising [1], [2], [3] separately but a handful of work has been done to achieve both simultaneously [9]. Recently Ankit.et.al [13] has use GAN based architecture for super-resolution to get super-resolved text to boost OCR performance. In [9] authors have proposed joint optimisation of two CNN based de-noising and SR module to improve OCR. These above frameworks lack in generalisation capability in terms of different scripts,printed/handwritten, font-styles etc. In all these frameworks, enhancement and recognition being two different modules in cascaded manner are unable to refine enhancement module weights to get more precise recognition results.

III. Contribution

Framework that jointly optimizes image-denoising and super-resolution of degraded document images through a coupled end-to-end trainable GAN framework. Use of cycle consistency loss [14] to refine HR target document image. Joint optimization of document image enhancement and recognition for degraded document images through a end-to-end robust OCR framework to boost recognition accuracy. Exhaustive experimentation with different types of data (printed and handwritten) in different scripts to show efficacy of proposed network and comparison with state-of-art frameworks for document image enhancement and recognition.

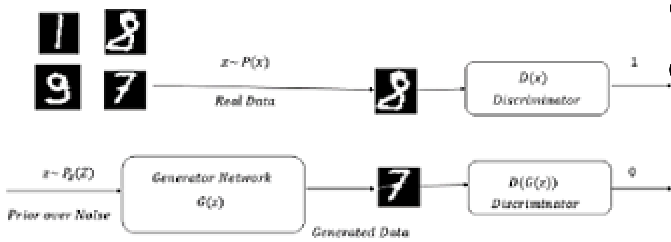
A. *Proposed Framework*

The Structure of proposed convolutional auto encoders



The proposed end-to-end network architecture for joint optimization of document image enhancement and recognition is shown in above figure. It comprises of two modules i.e., enhancement module and BLSTM-CTC based OCR module.

The Structure of proposed convolutional auto encoders



Enhancement module constitute of generative adversarial network (GAN) based fully convolutional REDNet [15] as generator G1 in network 1 and DBPN [16] as generator G2 in network 2 for degraded document image denoising and super-resolution respectively shown in Figure 2. The DBPN generator G3 is required for the down scaling of HR image to generate cyclic reconstructed image and calculate the cyclic-consistency loss [14] between low resolution denoised image and cyclic reconstructed image. This cyclic-consistency loss is then used for updating weights of enhancement module to additionally refine the results. The discriminators D1 and D2 used in network 1 and network 2 are similar to discriminator used in [10]. Both GAN based networks 1 and 2 are then trained in end-to-end manner for joint optimization of denoising and super-resolution. Bidirectional Long Short Term Memory (BLSTM) has efficiently overcome the constraints of traditional RNNs like diminishing gradient and requirement of pre-segmented data and has the competence to grab long range context [11]. LSTM utilizes multiplicative gates to trap the error and uphold the continual flow of error. This process is known as Constant Error Caraousal and aids to overcome the diminishing gradient issue. Graves et al. [12] has put forward a training technique called as CTC that was able to align sequential information and hence dodge the requirement of pre-segmented data. Using pre-trained weights of enhancement module both the enhancement module and BLSTM-CTC OCR module are trained in end-to-end configuration by CTC loss for joint optimization of document image enhancement and recognition.

IV. Data Sets

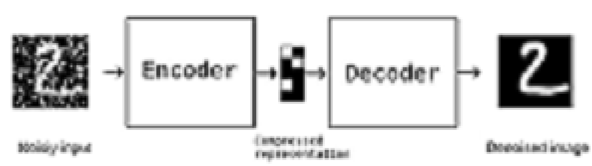
Proposed framework is validated on several datasets in different languages and compare them with different state-of-art algorithms: Publicly available handwritten datasets in English. We also trained and tested our systems on three Indic script datasets with different challenges: Devanagari (Hindi),English and Telugu. These datasets comprises of noisy and old degraded documents along with new printed books, accordingly offering a diverse set of recognition challenges. In order to make the system robust to low-resolution document images, we down sampled the high resolution images and used it as input to our system. All the documents are bilingual with English as the second script, and there are few pages which comprise of handwritten paragraphs as well making segmentation tough task. Before providing image as input to the framework skew correction and text graphics segmentation is performed. These datasets are splitted as: training (60Alongside the noise already present in the image, Gaussian [G (variance: 0.01 and 0.1)] and salt-and pepper [(SnP density: 0.01 and 0.1)] noises are added to LR noisy image.

V. Methodology

A. *Image Denoising*

Denoising is achieved through network 1 which constitute of the residual encoder decoder network (RED- Net) [15] as generator G1. Discriminator D1 is similar to discriminator used in [10]. The RED-Net is a fully convolutional network which primarily comprises of encoder (convolution), decoder (deconvolution) and skip connections. The encoder is used for feature extraction, which retains the essential objects in the image and simultaneously removes the corruptions.

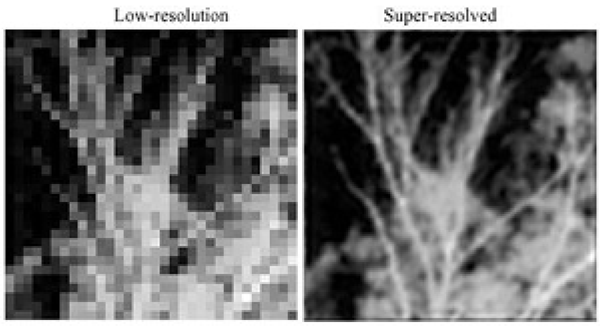
The Structure of proposed convolutional auto encoders



In decoder the deconvolutional layers are then summed up to restore the details. The output of decoder is a denoised (clean) version of input image. Skip connections are provided symmetrically between convolutional and deconvolutional layers because firstly, skip connections carry much image details and secondly, it aids on while back propagating the gradient to bottom layers making training of deeper network easier.

B. *Image SuperResolution*

superresolution



For the document image super-resolution in network 2 the DBPN [16] network is used as generator G2. Discriminator D2 is similar to discriminator used in [10]. The DBPN network comprises of three modules namely initial feature extraction, projection and reconstruction. The initial feature extraction unit is responsible for the construction of initial LR feature maps using convolutional blocks. Back projection stages is a array of projection units, switching between development of LR and HR feature maps in a way that every unit has access to the the yields of every single past unit. At last the reconstruction module results in the generation of target HR image using convolutional block. The network is optimized for three different losses naming, content loss i.e., Mean Square Error (MSE), adversarial loss and cyclic consistency loss.

C. *Joint Optimization of Denoising and Super Resolution*

We initialise the end-to-end coupled network for joint optimization of denoising and super-resolution of degraded document images by taking pre-trained weights of network 1 and network 2 as shown in Fig. 2. We consider pre-trained weights as initial weights to fine tune the combined network into an end-to-end manner.

D. *End-to-End Optimization for Degraded Document Image Enhancement and Recognition*

The pre-trained weights of enhancement module are saved and robust super resolved features from DBPNs last up projection layer are passed onto BLSTM having CTC as last layer for recognition of textual sequences. We then initialize the end-to-end network for joint optimization of enhancement module and BLSTM-CTC OCR module by taking pre-trained weights of enhancement module. We then fine tune the combined network of enhancement module and BLSTM-CTC OCR module in end-to-end configuration.

VI. Experimental Results

Enhancement models	Noise	English (Accuracy)		Hindi (Accuracy)		Telugu (Accuracy)		Handwritten (Accuracy)	
		OCR	PSNR	OCR	PSNR	OCR	PSNR	OCR	PSNR
HR Image	-	98.5		95.6		94.3		91.2	
Low Resolution Image	G 0.01	46.25		31.22		39.69		27.79	
	G 0.1	39.51		27.95		23.23		25.99	
	SnP 0.01	43.66		39.82		49.56		29.2	
	SnP 0.1	25.89		20.41		30.99		22.34	
SR for LR Image	-	96.21	49.2	95.4	45.47	92.17	46.21	89.98	40.22
LR 2X	G 0.01	29.87		20.23		19.34		19.12	
	G 0.1	20.91		15.18		16.87		17.3	
	SnP 0.01	36.2		22.84		28.90		18.97	
	SnP 0.1	19.99		14.83		21.19		15.56	
SR for LR 2X	-	96.78	34.2	95.13	33.5	93.49	32.67	89.99	31.13
LR 3X	G 0.01	15.29		12.86		10.23		10.01	
	G 0.1	10.45		9.98		9.67		9.12	
	SnP 0.01	13.85		10.57		10.19		11.01	
	SnP 0.1	9.98		8.45		8.28		8.23	
SR for LR 3X	-	94.33	19.56	93.1	18.87	90.51	17.34	86.19	15.89
LR 4X	G 0.01	4.9		3.0		2.98		1.99	
	G 0.1	3.12		2.89		1.45		0.87	
	SnP 0.01	3.87		3.76		2.56		1.67	
	SnP 0.1	1.90		1.84		1.23		0.45	
SR for LR 4X		93.34	10.23	92.12	9.45	89.9	9.12	85.7	8.99

As shown in the above table the accuracy is very high for high resolution images.For low resolution images the accuracy is very less and it is improved by super resolution.

VII. Conclusion

Exhaustive comparison of our model with other state-of-the-art models have been depicted in our paper. End-to-End trainable network that jointly optimizes image enhancement and text recognition would make a high-performing robust OCR system. Our model performed exceptionally well and showed the robust and efficient performance over printed and handwritten low-resolution de-graded document images. We have achieved the state-of-the-art result over both printed and handwritten datasets for English as well as three Indic Scripts.

VIII. References

1.Banerjee, A. M. Namboodiri, and C. V. Jawahar, "Contextual restoration of severely degraded document images," in IEEE Computer Vision and Pattern Recognition, 2009.

2.Prateek Sarkar, H. S. Baird, and Xiaohu Zhang, "Training on severely degraded text-line images," in International Conference on Document Analysis and Recognition(ICDAR), 2003.

3.M. D. Gupta, S. Rajaram, N. Petrovic, and T. S. Huang, "Restoration and recognition in a loop," in IEEE Computer Vision and Pattern Recognition (CVPR), June 2005.

4.Dmitry Datsenko and Michael Elad, "Example-based single document image super-resolution: a global map approach with outlier rejection," Multidimensional Systems and Signal Processing, 2007.

5.Chao Dong, Ximei Zhu, Yubin Deng, Chen Change Loy, and Yu Qiao, "Boosting optical character recognition: A superresolution approach," CoRR, vol. abs/1506.02211, 2015.

6.K. Donaldson and G. K. Myers, “Bayesian super-resolution of text in video with a text-specific bimodal prior,” in IEEE Computer Vision and Pattern Recognition (CVPR), 2005.

7.H. Zhang, D. Liu, and Z. Xiong, “Cnn-based text image super-resolution tailored for ocr,” in 2017 IEEE Visual Communications and Image Processing (VCIP), 2017.

8.Hiep Q. Luong and Wilfried Philips, “Robust reconstruction of low-resolution document images by exploiting repetitive character behaviour,” International Journal of Document Analysis and Recognition (IJDAR), 2008.

9.M. Sharma, A. Ray, S. Chaudhury, and B. Lall, “A noise resilient super-resolution framework to boost ocr performance,” in 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), 2017.

10.Christian Ledig, Lucas Theis, Ferenc Huszar, Jose Caballero, Andrew P. Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe Shi, “Photo-realistic single image superresolution using a generative adversarial network,” CoRR, vol. abs/1609.04802, 2016.

11.A. Graves, M. Liwicki, S. Fernndez, R. Bertolami, H. Bunke, and J. Schmidhuber, “A novel connectionist system for unconstrained handwriting recognition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009.

12.Alex Graves, Santiago Fernandez, Faustino Gomez, and Jurgen Schmidhuber, “Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks,” in Proceedings of the 23rd International Conference on Machine Learning, 2006, ICML ’06.

13.Ankit Lal and C V. Jawahar, “Enhancing ocr accuracy with super resolution,” in ICPR 2018, 08 2018, pp. 3162–3167.

14.Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, “Unpaired image-to-image translation using cycleconsistent adversarial networks,” CoRR, vol. abs/1703.10593, 2017.

15.Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang, “Image restoration using convolutional auto-encoders with symmetric skip connections,” CoRR, vol. abs/1606.08921, 2016.

16.Muhammad Haris, Greg Shakhnarovich, and Norimichi Ukita, “Deep back-projection networks for super-resolution,” CoRR, vol. abs/1803.02735, 2018.

17.Urs-Viktor Marti and H Bunke, “The iam-database: An english sentence database for offline handwriting recognition,” International Journal on Document Analysis and Recognition, vol. 5, pp. 39–46, 11 2002.

18.R. Smith, “An overview of the tesseract ocr engine,” in International Conference on Document Analysis and Recognition, 2007, ICDAR.