A survey of Deep Learning Techniques uses GAN and CNN for the Pre-Processing and Enhancement of Historical Documents

Dharmen B. Shah
Assistant Professor,
Department of Computer Science,
VNSGU, Surat – 395007, Gujarat, India
Email:dharmenbshah@gmail.com

Dr. Apurva A. Desai
Professor & Head,
Department of Computer Science,
VNSGU, Surat – 395007, Gujarat, India
Email:aadesai@vnsgu.ac.in

Abstract:

We have a rich collection of information and knowledge preservedover generations in historical documents—manuscripts. Their preservation is essential to pass on the knowledge from one generation to another. One way to preserve this knowledge source is through the digitization. As these documents are maintained for a long time, they are susceptible to degradation from factors such as ageing, ink fading, stains, bleed-through, annotations, and rough handling. These issues will create additional challenges forresearchers working with digitized versions of these documents, particularly in the context of automated analysis and recognition. That is why Document Image enhancement becomes a critical pre-processing step for accuracy in further processing. Various traditional image processing techniques can be useful in preprocessing, but fail to give satisfactory results, especially while handling the complex and diverse nature of document degradation. New approaches have emerged with the advancement of machine learning and deep learning that demonstrate superior results in enhancing degraded documents. This paper provides a detailed review of GAN (Generative Adversarial Network) and CNN (Convolutional Neural Network) based methods applied to the six major document enhancement tasks: binarization, deblurring, denoising, defading, watermark removal, and shadow removal.The study can help to gain knowledge about the challenges of historical document preprocessing and use it to develop more reliable and efficient solutions to the problem.

1. Introduction

Ancient manuscripts contain rare information about our heritage and culture, and the digitization of these valuable documents is important for making them available to the next generation. Among all the document types, ancient documents/manuscripts of machine printed or handwritten types are very highly affected by noise [1]. Those historical documents are often degraded due to watermarks, stamps, ink stains, bleed-through, poor storage, humidity and environmental factors, ageing, etc. Historical document images with degradation have low visual quality and readability, which makes automatic document analysis tasks like handwritten character recognition (HCR) very challenging. Those historical manuscripts are often in very large numbers, which makes it practically infeasible to manually enhance each image; thus, it is essential to develop methods that can automatically enhance the document images to improve their readability and restore the missing or corrupted information.

Researchers have been working on the problem of document image enhancement for a long time, and some good results have been achieved by many researchers. The use of machine learning is increasing in various digital image processing tasks, such as object detection [3],

dataset creation [2] and image enhancement [4]. It has been shown that such deep learning-based methods achieve good results and outperform the traditional methods. In the same way, deep learning based methods for document image enhancement have created great interest in recent years. The goal of this survey is to review these methods and discuss their features, advantages, limitations and identify the opportunities in future research.

2. Documents' life cycle and degradation

We must understand the various types of degradation before rectifying them through an appropriate preprocessing procedure. Historical document images may be degraded at different stages, i.e. at the time of creation, during their utilization, digitization and processing.

1) Degradation during creation

Those historical manuscripts are written using different types of writing materials like Tamra Patra, palm leaves, cloths, papers, etc. by the document's writers. Degradation at the time of creation occurs mostly due to the medium's quality and the individual's writing style.

2) Degradation during utilization

Once the document is created, it is used as and when required. Means that they may be kept aside for referencing purposes or may be used too frequently at times. In this case, the documents are subject to external degradation due to humans or the environment. Noise created due to the usage of the document and its ageing is called external degradation or physical noise. Some examples are,

i. The quality of the material on which those manuscripts are written is getting reduced.
ii. The document may suffer from stains, scratches or cracks due to frequent usage.
iii. Annotations like comments, explanations and stamps may be included.

3) Degradation during digitization

For digital image processing, these historic documents must be converted to digital form, the process called digitization. This is done through devices like scanners and cameras. Some degradation arises in the process of digitization due to

i. Improper document alignment causes the skew.
ii. Pixel sensitivity variation, which leads to resolution variation in the document image.
iii. Vibration or shakiness in the camera device can cause a blurred image.

4) Degradation during processing

This refers to all types of degradation that occur after the document image is created. This involves the degradation caused due to,

i. Use of lossy compression, like JPEG, will cause some degradation.

ii. The nature of the transmission medium will also cause some amount of degradation.

iii. Use of an inappropriate binarization technique will add noise in the binary image.

3. Document Image Enhancement

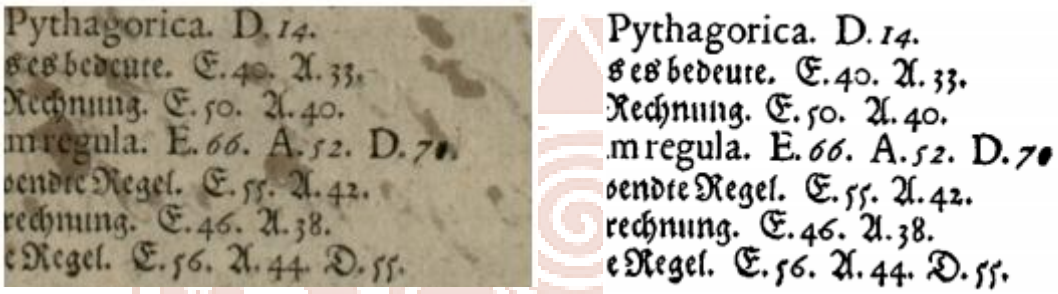
After the digitization of the historic document, we must do the preprocessing of the document image to make it ready for digital image processing. One important phase of the preprocessing is the image enhancement. There are several tasks that can be done related to the document image enhancement, like

Figure 1: Sampleimages of degraded historical document images.

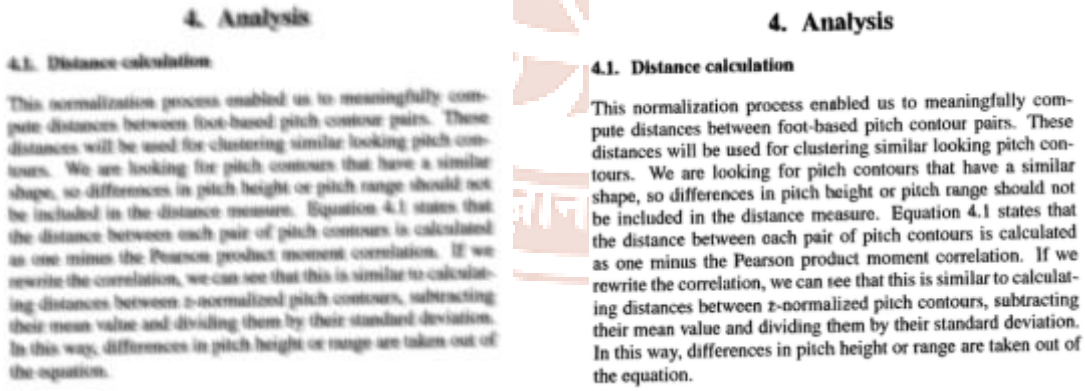
1) **Binarization**: The purpose of binarization is to separate background from foreground, i.e. text, to remove noise, ink stain, bleed-through, wrinkles, etc. The output will be a binary image.

2) **Deblur**: this task aims to remove different types of blurs like Gaussian, motion, de-focus, etc.
3) **Denoise**: Denoising will remove various types of noise like salt and paper, wrinkles, dog-eared, background stain, etc., from the document image.
4) **Defade**: This step is to improve a faded document image.A document image can be lightly or heavily faded over a period of time due to overexposure, sunlight, etc.
5) **Watermark removal**: Sometimes, historical documents contain the stamps or marks of the institute they belong to or some other types of stamps, which make the text under it unreadable. This task aims to remove such watermarks.
6) **Shadow removal**: While capturing the document image, if care is not taken, the shadow of the light source may be added to the captured document image. This task will remove this type of effect.

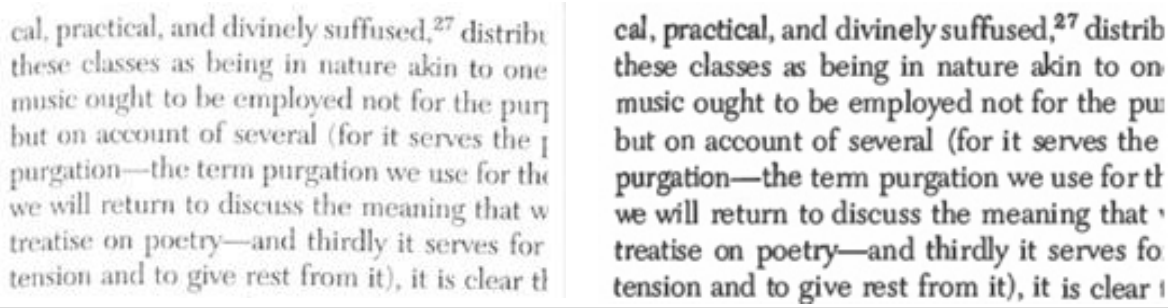
1) Binarization task



2) Deblur task



3) Defading task



4) Denoise task

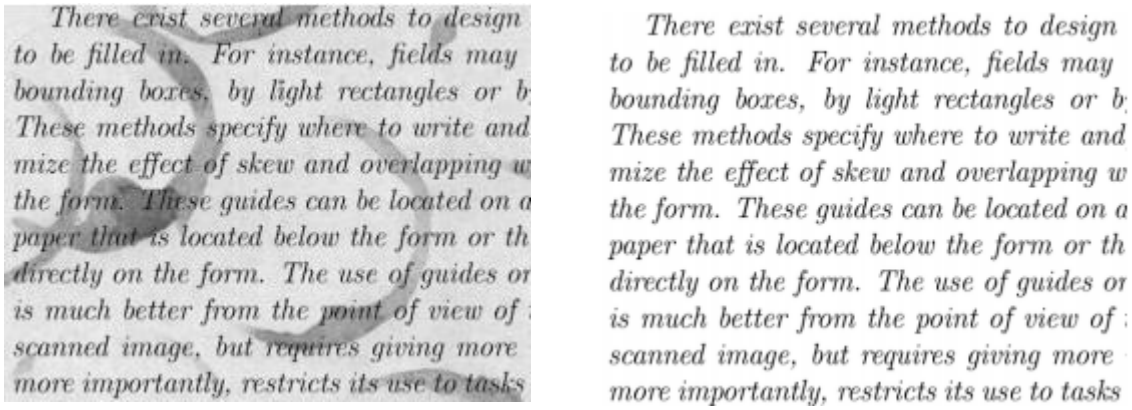
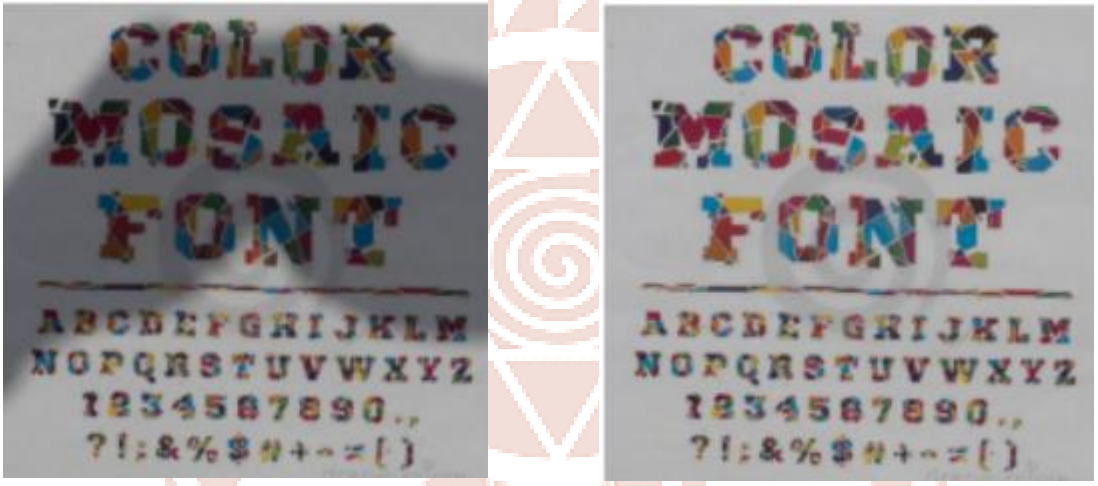
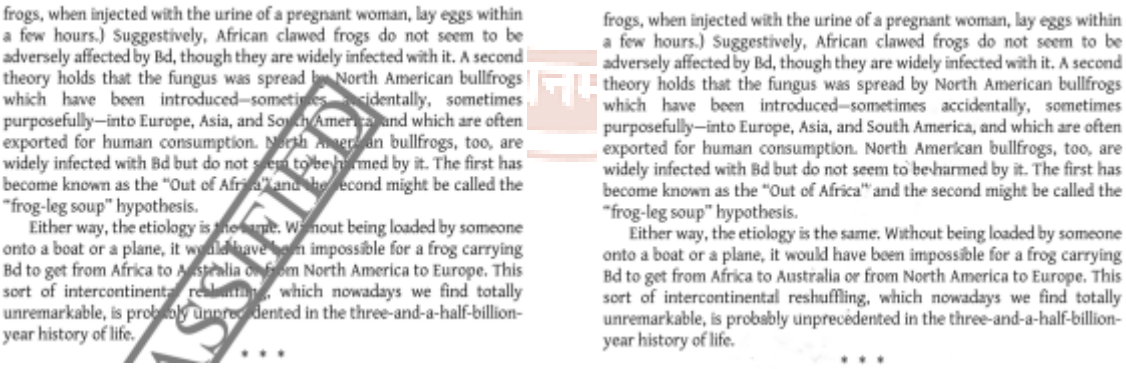


Figure 2: Sample images for different document image enhancement tasks. The image on the left is the input, and theright image is the clean/ground truth image.

5) Shadow Removal task



6) Watermark Removal task

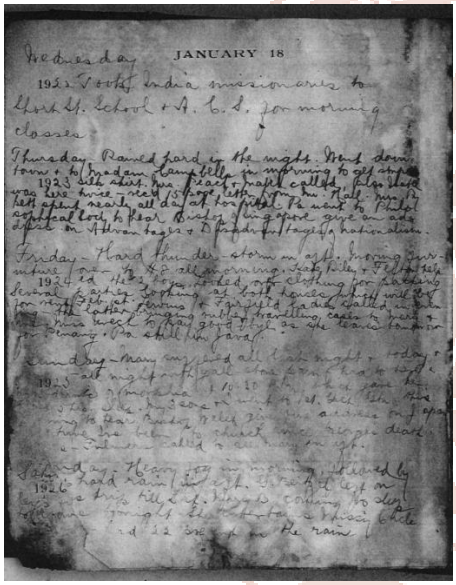


4. Datasets

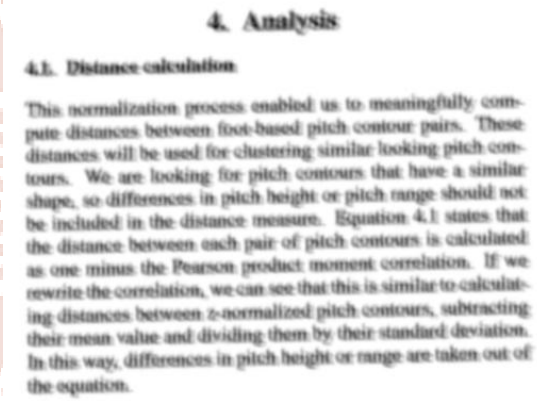
In this section, we will discuss the various datasets used in the literature to perform various image enhancement tasks. The description of the dataset is given below, and Table 1 gives a summary of these datasets. Also, through Figure 3, we have shown some sample images of the document from these datasets.

Dataset	Task	No of Images	Resolution(Pixels)	Real vsSynthetic
Bickley diary [18]	Binarization	7	1050 x 1350	Real
NoisyOffice [19]	Denoising	288	variable	Real/Synthetic
S-MS [20]	Multiple	240	1001 x 330	Synthetic
Tobacco 800 [21]	Denoising	1290	(1200x1600) - (2500x3200)	Real
DIBCO'17 [24]	Binarization	10	(1050x608) - (2092x951)	Real
H-DIBCO'17 [25]	Binarization	10	(351x292) - (2439x1229)	Real

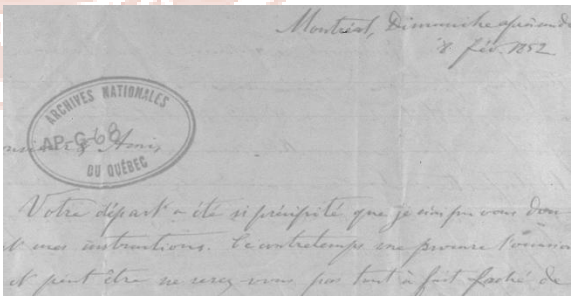
Table 1: Specifications of the datasets used for different document image enhancement tasks.



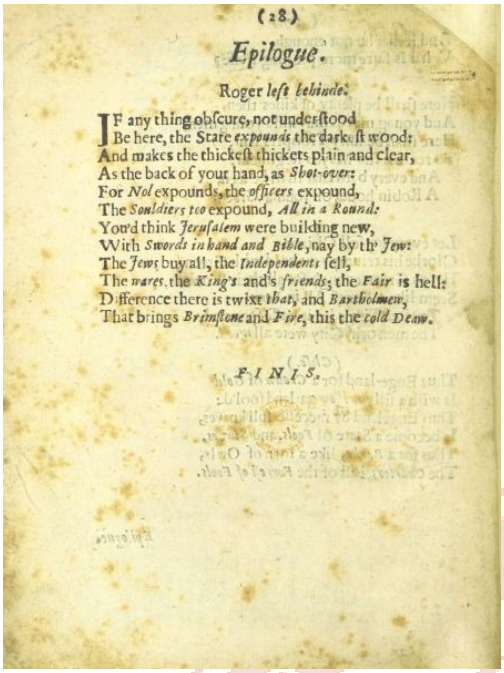
(a) Sample image from Bickley Diary Dataset



(b) Sample image from BMVC Dataset



(d) Sample image from S-MS Dataset


(28)

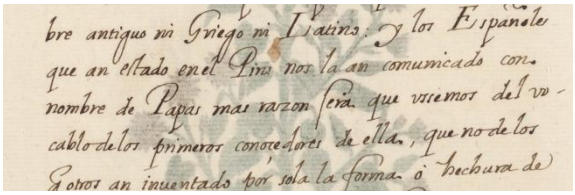
Epilogue.

Roger *left behinde:*

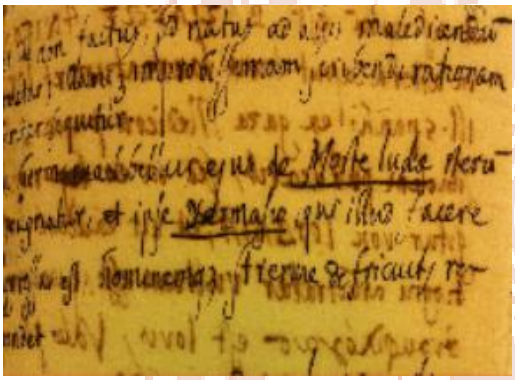
IF any thing obſcure, not underſtood
Be here, the State *expounds* the darkeſt wood:
And makes the thickeſt thickets plain and clear,
As the back of your hand, as *Shot-over:*
For *Nol* expounds, the *officers* expound,
The *Souldiers too* expound, *All in a Round:*
You'd think *Jeruſalem* were building new,
With *Swords in hand and Bible*, nay by th' *Jew:*
The *Jews* buy all, the *Independents* ſell,
The *wares* the *King's* and's *friends*; the *Fair* is hell:
Difference there is twixt *that*, and *Bartholmew*,
That brings *Brimſtone* and *Fire*, this the *cold Dew.*

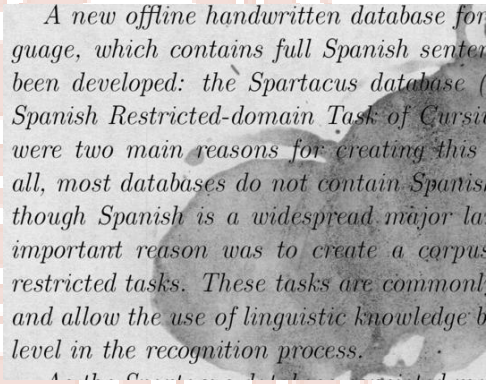
FINIS.

(c) Sample image from DIBCO Dataset


bre antiguo ni Griego ni Latino. y los Espanoles
que an estado en el Piru nos la an comunicado con
nombre de Papas mas razon sera que vsemos del vo-
cablo de los primeros conoscedores de ella, que no de los
Gotos an inventado por sola la forma ó hechura de

(e) Sample image from H-DIBCO Dataset


(f) Sample image from the MCS dataset


A new offline handwritten database for
guage, which contains full Spanish senter
been developed: the Spartacus database (
Spanish Restricted-domain Task of Cursi
were two main reasons for creating this
all, most databases do not contain Spanis
though Spanish is a widespread major la
important reason was to create a corpus
restricted tasks. These tasks are commonl
and allow the use of linguistic knowledge b
level in the recognition process.

(g) Sample image from Noisy Office Dataset

Figure 3: Sample images from datasets for document image enhancement tasks.

- **Bickley diary**: The images of the Bickley diary dataset are taken from a photocopy of a diary written about 100 years ago. These images suffer from different kinds of degradation, such as water stains, ink bleed-through, and significant foreground text intensity. This dataset contains 7 document images/pages along with the binarized/clean ground truth images. [18]

- **NoisyOffice:** This dataset contains two sets of images: 1) Real Noisy Office: it contains 72 grayscale images of scanned noisy images, 2) Simulated Noisy Office: it contains 72 grayscale images of scanned simulated noisy images for training, validation and test.[19]

- **S-MS:** SynchromediaMultiSpectral Ancient document: Multi-spectral imaging (MSI) represents an innovative and non-destructive technique for the analysis of materials

such as ancient documents. They collected a database of multispectral images of ancient handwritten letters. This database consists of multispectral images of 30 real historical handwritten letters. These extremely old documents were all written by iron gall ink and date from the 17th to the 20th century.[20]

- **Tobacco 800:** This is a publicly available subset of 42 million pages of documents that are scanned with various equipment. It contains real-world documents, and it contains different types of noise and artefacts, such as stamps, handwritten texts, and ruling lines, on the signatures. [21]

- **DIBCO and H-DIBCO:** These datasets were introduced for the Document Image Binarization Contest in 2009. They are DIBCO 2009[24] andH0DIBCO 2010[25]. DIBCO datasets contain both printed and handwritten document images, mainly for the binarization task.

- **Blurry document images (BMVC):**The training data contains 3M train and 35k validation 300x300 image patches. Each patch is extracted from a different document page, and each blur kernel used is unique.[22]

- **Monk Cuper Set (MSC):** This dataset contains 25 pages sampled from real historical documents, which are collected from the Cuper book collection of the Monk system. MSC documents suffer from heavy bleed-through degradations and textural background.[23]

5. Metrics

The evaluation metrics that are used for different document image enhancement tasks.

- **Peak signal-to-noise ratio (PSNR)**

  This metric is reference-based, provides pixel-wise evaluation, and can indicate the effectiveness of document enhancement methods in terms of visual quality. It measures the ratio between the maximum possible value of a signal and the power of the distorting noise that affects its quality of representation. In other words, it measures the closeness of two images. A high value PSNR indicates greater similarity between the two images. MAX is the pixel value that is possibly maximum. When the pixel size is 8 bits per sample, it is 255. Given two MxNimages, this metric would be formulated as follows:

$$PSNR = 10\log(\frac{MAX^2}{MSE}) \quad (1)$$

Where

$$MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - I'(x,y))^2}{MN} \quad (2)$$

- **Character Error Rate (CER):** CER is computed based on the Levenshtein distance. CER is calculated in terms of the minimum number of operations to be performed at the character level to transform ground truth text into the OCR output text. The CER formula is as follows,

$$CER = \frac{S + D + I}{N} \quad (3)$$

Where S is the Substitutions, D is the Deletions, I isthe Insertions, and N is the character count in the reference or ground truth text.

CER represents the percentage of characters in the reference text that were incorrectly predicted or misrecognized in the OCR output. The lower the CER value, the better the performance of the OCR model. Normalized CER can ensure that it will not fall out of the 0-100 range due to too many insertions. In normalized CER, C is the number of correct recognitions. Normalized CER is formulated as follows,

$$CER_{normalized} = \frac{S + D + I}{S + D + I + C} \quad (4)$$

- **Word Error Rate (WER):** The OCR performance can be evaluated using WER on paragraphs and words. The computation of WER is similar to CER, but WER works at the word level.It represents the number of word substitutions, deletions, or insertions needed to transform one sentence into another. WER is formulated below,

$$WER = \frac{S_w + D_w + I_w}{N} \quad (5)$$

- **F-measure:** The precision and recall are used to compute the harmonic mean known as the F-measure score. Where the positive predictive value is called precision, and the sensitivity in diagnostic binary classification is called recall.F-measure is formulated as follows,

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \quad (6)$$

Where,

$$Recall = \frac{TP}{TP + FN} \tag{7}$$

$$Precision = \frac{TP}{TP + FP} \tag{8}$$

Where TP is true positive, FP is false positive, and FN is false negative.

- **Distance Reciprocal Distortion Metric (DRD):** The visual distortion in binary images is measured using DRD. It properly correlates with the human visual perception, and it measures the distortion for all pixels as follows,

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{9}$$

where NUBN is the number of the non-uniform (not all black or white pixels) 8x8 blocks in the ground truth image, and the distortion of the kth flipped pixel that is calculated using a 5x5 normalized weight matrix WNm is represented by DRDk. The weighted sum of the pixels in the 5x5 block of the ground truth image differs from the centre kth flipped pixel at(x, y) in the binarization result image represented as DRDk (equation 10)

$$DRD_k = \sum_{i=-2}^{2} \sum_{j=-2}^{2} |GT_k(i,j) - B_k(x,y)| \times W_{Nm}(i,j) \tag{10}$$

6. Document Image Enhancement Methods

Deep Learning has recently demonstrated impressive results in a number of document image enhancement domains. The image enhancement algorithms that we have discussed and described here are mainly based on two kinds of deep learning techniques that are widely used and have provided the highest accuracy when compared to other algorithms.

A framework comprising two competing neural networks: a generator and a discriminator, is known as a GAN (Generative Adversarial Network). The generator creates new, realistic data (e.g., images or text) from random noise, while the discriminator tries to distinguish it from real training data. Through this adversarial process, the generator learns to produce increasingly convincing outputs that fool the discriminator, thereby generating high-quality, authentic data.

CNN (Convolutional Neural Network) is a deep learning algorithm that automatically learns to recognize visual patterns in images and videos by processing grid-like data through layers of filters. CNNs are fundamental to computer vision

applications, enabling tasks such as object recognition, image classification, and medical image analysis.

Hradis et al. [32] introduced a method for the document image deblurring problem. The authors proposed a 15-layerconvolutional neural network model to deblur images without assuming any priors, for example,any specific blur ornoise models. Types of blurs focused on are a combination of realistic defocus blur and camera shakeblur. The proposed networkclaims to significantly outperform existing blind deconvolution methods, including those optimized for text, both in terms of image quality and OCR accuracy.

Xu et al.[34] also proposed a deblurring algorithm to directly restore a high-resolution deblurred image from a blurry input. A multi-class GAN model was developed to learn a category-specific prior and process multi-classimage restoration tasks, using a single generator network.The authors employed a deep CNN architecture proposed byHradis et al. [32] in an adversarial setting. The algorithm contains up-sampling layers, which are fractionally-stride convolutional layers, aka deconvolution layers. Since their model has a discriminator network in addition to the generator network, it ismore complex and has more parameters compared to [32].The quality of the generated images was evaluated in terms of PSNR and SSIM, but the deblurred document images were not evaluated in terms of OCR performance, and no CER/WER is reported.

Tensmeyer et al. [39] focused on the degraded historical manuscript images binarization, and formulatedbinarization as a pixel classification learning task. They have developed a Fully Convolutional Network (FCN) that operates on multiple full-resolution images. As they claim, the proposed technique can also be applied to other domains, such as palm leaf manuscripts, with good results.

Zhao et. al. [27] work on the denoising and deblurring problems and proposed a method for document imagerestoration called Skip-Connected Deep Convolutional Autoencoder (SCDCA), which is based on residual learning.They employed two types of skip connections: identity mapping between convolution layers inspired by residual blocks, and another type that connects the input to the output directly. These connections assist the networkto learn the residual content between the noisy and clean images instead of learning an ordinary transformationfunction. The proposed network was inspired by Hradis et al. [32],which is a 15-layer CNN.

Sharma et al. [29] cast the image restoration problem as an image-to-image translation task, i.e, translating a document with noise like background noise, blurred, faded, or watermarked, etc., to a clean document using a GAN approach. To do so, they employed a CycleGAN model, which is an unpaired image-to-image translation network,for cleaning the noisy documents. They also synthetically created a document dataset for watermark removal anddefading problems by inserting logos as watermarks and applying fading techniques on the Google News dataset ofdocuments.

Gangeh et al. [28] proposed an end-to-end document enhancement pipeline that takes in blurry andwatermarked document images and produces clean documents. They trained an auto-encoder model that works ondifferent noise levels of documents. They have adopted the neural network architecture described by Mao et al. [40] called REDNET and designed aREDNET with 15 convolutional layers and 15 deconvolutional layers, including 8 symmetric skip connections betweenalternate convolutional layers and the mirrored deconvolutional layers. The advantage of this method compared to a fully convolutional network is that pooling and unpooling, which tend to eliminate image details, is avoided for low-level image tasks such as image restoration, resulting in higher resolution outputs. The key differences of this work from [27]are the use of a larger dataset and training a blind model.

Souibgui et al. [30] proposed an end-to-end framework called Document Enhancement Generative Adversarial Networks (DE–GAN). This network is based on conditional GANs -cGANs, a network to restore severely degradeddocument images. The tasks that are studied in this paper are document clean up, binarization, deblurring and watermarkremoval. Due to the unavailability of a dataset for the watermark removal task, the authors synthetically created a watermarkdataset including the watermarked images and their clean ground truth.

Lin et al. [38] proposed the Background Estimation Document Shadow Removal Network (BEDSR-Net),which is the first deep network designed for document image shadow removal. They designed a background estimationmodule to extract the document's global background colour. During the process of estimating the backgroundcolor, this module learns information about the spatial distribution of background and also the non-background pixels.They created an attention map through encoding this information. Having estimated the global background color andthe attention map, the shadow removal network can now effectively recover the shadow-free document image.

BEDSR-Net can fail in some situations, including when there is no single dominant color, such as a paper entirely witha color gradient, and another case is when the document is entirely shadowed, or multiple shadows are formed bymultiple light sources.

Table 2 shows the summary descriptions of methods we have reviewed in this paper and the task and document types handled by those methods, whereas

Table 3 summarises the advantages, disadvantages and results of those methods.

Methods	**Document Image Enhancement Tasks**						**Document Type (Hand Written / Printed)**	**Method**
	Binarization	Deblur	Denoise	Defade	Watermark Removal	Shadow Removal		
Gangeh et al. [26]	-	✓	✓	✓	✓	-	-	GAN
Zhao et al. [27]	-	✓	✓	-	-	-	-	CNN
Sharma et al. [29]	-	✓	-	✓	✓	-	Printed	GAN
Sharma et al. [29]	-	-	-	-	-	✓	Handwritten	GAN
Souibgui et al. [30]	-	✓	-	-	✓	-	Both	GAN
Gangeh et al. [28]	-	✓	-	-	✓	-	Printed	CNN
Hradiš et al. [32]	-	✓	-	-	-	-	-	CNN
Jemni et al. [33]	✓	-	-	-	-	-	Handwritten	GAN
Xu et al. [34]	✓	-	-	-	-	-	Printed	GAN
Souibgui et al. [31]	-	✓	-	-	-	✓	Printed	GAN
Calvo-Zaragoza et al. [35]	✓	-	-	-	-	-	Both	CNN
Dey et al. [36]	✓	-	✓	-	-	-	Printed	CNN
Li et al. [37]	✓	-	-	-	-	-	Both	CNN

Table 2: Details of the methods reviewed in this paper.

Souibgui et al. [31] focused on documents that are digitized using smartphone cameras. They statedthat these types of digitized documents are highly vulnerable to capturing various distortions including but not limitedto perspective angle, shadow, blur, warping, etc. The authors proposed a conditional generative adversarial networkthat maps the distorted images from its domain into a readable domain. This model integrates a recognizer in thediscriminator part for better distinguishing the generated document images.

Gangeh et al. [26] studied an end-to-end unsupervised deep learning model to remove multiple types of noise, including salt & pepper noise, blurred and/or faded text, and watermarks from documents. In particular, they proposed a unified architecture by integrating a deep mixture of experts proposed by Wang et al. [41] with

a cycle-consistent GAN as the base network for the document image blind denoising problem.

Dey et al. [36] work on the document image cleanup problem on embedded applications such as smartphone apps, which usually have memory, energy, and latency limitations. They proposed a light-weight encoder-decoder CNNarchitecture, incorporated with perceptual loss. They proved that in terms of the number of parameters and product-sumoperations, their models are 65-1030 and 3-27 times, respectively, smaller than existing SOTA document enhancementmodels.

Jemni et al. [33] focused on enhancing handwritten documents and proposed an end-to-end GAN-basedarchitecture to recover the degraded documents. The proposed architecture integrates a handwritten text recognizer, which makes the generated binary document image more legible. The approach claims to be the first work that uses the text information while binarizing handwritten documents, according to the authors.They performed experiments on degraded Arabic and Latin handwritten documents and showed that their modelimproves both the visual quality and the legibility of the degraded document images.

Li et al. [37] proposed a document binarization method called SauvolaNet. They investigated theclassic Sauvola [42] document binarization method from a deep learning perspective and proposed a multi-windowSauvola model. They also introduced an attention mechanism to automatically estimate the required Sauvola windowsizes for each pixel location, therefore could effectively estimate the Sauvola threshold.The proposed network has three modules: Multi-Window Sauvola, Pixelwise Window Attention, and Adaptive SauvolaThreshold. The Multi-Window Sauvola module reflects the classic Sauvola but with trainable parameters and multi-window settings. The next module, which is Pixelwise Window Attention, is in charge of estimating the preferred windowsizes for each pixel. The other module, Adaptive Sauvolva Threshold, combines the outputs from the other two modulesand predicts the final adaptive threshold for each pixel. The SauvolaNet model significantly reduces the number ofrequired network parameters and achieves SOTA performance for the document binarization task.

Methods	Advantages	Disadvantages	Results
Gangeh et al.[26]	- Handles multiple noises, including salt and pepper noise, faded, blurred, and watermarked documents, in an end-to-end manner. - It does not rely on a paired document image	- Computationally complex.	- Method has the best results in terms of PSNR and OCR
Lin et al. *[38]*	- First deep learning-based approach for shadow removal. - It works on both grey-scale and RGB images	- Computationally complex. - It does not work well on images with complex backgrounds and layouts. - It works well on partially shadowed documents only	- It achieves the best results in terms of PSNR/SSIM compared to Four previous works when evaluated on five different datasets. - It also generalizes relatively well on real-world images
Zhao et al. [27]	- Method is fast and easy to implement	- Inadequate qualitative and quantitative results	- Marginal PSNR improvement
Gangeh et al.[28]	- Works on both grey-scale and RGB watermarks. - Works on blurry images withvarious intensities.	- Inadequate quantitative evaluation and comparison with previous work	- Effectively removes watermark and blur. - Improved OCR on a small test set of nine images.
Souibgui et al. [30]	- Flexible architecture could be used for other document degradation problems. - First work on dense watermark and stamp removal problems. - Generalize well on real-world images. - Pre-trained models are publicly available.	- Computationally complex. - It needs a threshold to be pre-determined and needs to be tuned per image, which makes this method is less practical.	**Binarization:** Achieves best results in terms of PSNR, *F-measure*, FPSand DRD compared to the top five competitors. **Watermark:** Achieves the best results in terms of PSNR/SSIM compared to the three previous works. **Deblur:** Achieves the best results in terms of PSNR compared to the two previous works.
Sharma et al.[29]	- Adaptable for both paired and unpaired supervision scenarios	-	- Marginal improvement in terms of PSNR

Hradiš et al.[32]	- Small and computationally efficient network. - Can be used on mobile devices.	- Adds ringing artefacts in somesituations. - Does not work well on uncommon words when the image isseverely blurred.	- Outperforms other methods interms of PSNR and Character Error Rate compared to previous four work.
Xu et al. *[34]*	- Computationally efficient network. - It deblurs and super-resolves simultaneously.	- Does not generalize well forgeneric images. - OCR performance evaluation isignored, and only visual quality of the documents are evaluated.	- Performs favourably against previous work on both synthetic andreal-world datasets.
Souibgui et al. [31]	- It handles multiple camera distortions. - It incorporates a text recognizerfor generating more legible images.	- It handles multiple camera distortions. - It incorporates a text recognizerfor generating more legible images.	- Achieves best results in termsof Character Error Rate and second best in terms of PSNRcompared to the previous three work.

Table 3: Performance Comparison ofmethods reviewed in this paper.

7. Conclusion and Future Work

In this paper, we have reviewed deep learning-based methods for various document image enhancement tasks, especially with context to historical documents, focusing on tasks like binarization, deblurring, denoising, defading, watermark removal and shadow removal. We have summarised the datasets of historic documents which are being used for this task and also discussed about the matrices used to evaluate the performance of those methods. We have discussed each method with reference to its features, performance and challenges in the context of each task.

Document image enhancement tasks are still very challenging due to the variations in challenges and characteristics of the different types of historical document images; some of the enhancement tasks are either not studied at all or studied in a very limited context with various constraints. This gives huge opportunities to researchers to work in this area. Based on our study, we can summarise the following areas in which researchers can work ahead.

Overexposure and underexposure correction tasks

When we capture the document image using a camera, it is very likely that the light parameters are not ideal, which leads to overexposure or underexposure. Overexposure occurs when too much light is used while digitizing the document. And on the other

hand, underexposure occurs when the light condition is very poor. This problem is different from shadow removal. And work can be done to develop deep learning-based methods to solve these problems, at the same time, no public dataset is available for such images with their ground truth image, so someone can work on dataset creation as well.

Defading task

This task is the understudied task of document image enhancement, and current work considers only lightly faded documents, while in reality, historical documents are likely to be heavily faded. Those are very challenging for the OCR engine, and a deep learning-based model can be developed to work with heavily faded documents. Just like the previous problem, here also, no public dataset is available, so the work can be done for that as well.

Ghosting and bleed-through removal

Ghosting occurs when text from the other side of the page can be seen, but the ink does not completely come through to the other side. Bleed-through, on the other hand, occurs where ink seeps into the other side and interferes with the text on the other side. It is very frequently seen in the case of historical documents, which impacts the performance of the OCR system. Binarization methods are employed which can remove this type of degradation partially but are not evaluated in terms of CER. Therefore, more effective methods are required to be developed to remove these artefacts, which will eventually lead to improvement in the performance of the OCR system.

OCR performance evaluation

The objective of the document image enhancement is to improve the result of the OCR Systems in automated document analysis. Currently, we do not have the dataset of historical document images with their ground truth text to check the performance of the image enhancement methods in terms of OCR improvement. Current methods either ignore or test only a very few images in terms of OCR performance, which means a separate study is needed to collect the dataset and benchmark current methods against this benchmark dataset.

References

1. Gupta MR, Jacobson NP, Garcia EK (2007) OCR binarization and image pre-processing for searching historical documents. Pattern Recogn. 40(2):389–397

2. ANVARI, Z., AND ATHITSOS, V. A pipeline for automated face dataset creation from unlabelled images.In Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to AssistiveEnvironments (2019), pp. 227–235
3. REDMON, J., DIVVALA, S., GIRSHICK, R., AND FARHADI, A. You only look once: Unified, real-time objectdetection. In Proceedings of the IEEE conference on computer vision and pattern recognition (2016), pp. 779–788.
4. ANVARI, Z., AND ATHITSOS, V. Enhanced cyclegan dehazing network. In VISIGRAPP (4: VISAPP) (2021),pp. 193–202
5. Munish Kumar, Manish Kumar Jindal, Rajendra Kumar Sharma, and Simpel Rani Jindal. 2019. Character and numeral recognition fornon-Indic and Indic scripts: a survey. Artificial Intelligence Review 52, 4 (2019), 2235-2261.
6. Munish Kumar, M. K. Jindal, R. K. Sharma, and SimpelRaniJindal. 2018. Performance Comparison of Several Feature Selection Techniquesfor Online Handwritten Character Recognition. In Proc. of International Conference on Research in Intelligent and Computing in Engineering(RICE). IEEE, 1-6
7. Bappaditya Chakraborty, Bikash Shaw, Jayanta Aich, Ujjwal Bhattacharya, and Swapan Kumar Parui. 2018. Does Deeper Network Leadto Better Accuracy: A Case Study on Handwritten Devanagari Characters. In Proc. of 13th IAPR International Workshop on Document Analysis Systems (DAS). IEEE, 411-416.
8. B. Shaw, U. Bhattacharya, and S. K. Parui. 2015. Oline Handwritten Devanagari Word Recognition: Information Fusion at Feature and Classifier Levels. In Proc. of ACPR. IEEE, 720-724.
9. U. Bhattacharya, M. Shridhar, S. K. Parui, P. K. Sen, and B. B. Chaudhuri. 2012. Oline recognition of handwritten Bangla characters: an efficient two-stage approach. Pattern Analysis and Applications 15, 4 (2012), 445-458.
10. K. Mullick, S. Banerjee, and U. Bhattacharya. 2015. An efficient line segmentation approach for handwritten Bangla document image. InProc. of ICAPR. 1-6.
11. M. Kumar, M.K. Jindal, R.K. Sharma, et al. 2017. Oline Handwritten Gurmukhi Character Recognition: Analytical Study of DifferentTransformations. Proceedings of the National Academy of Sciences, India, Section A: Physical Sciences 87 (2017), 137-143.

12. M. Kumar, M.K. Jindal, R.K. Sharma, et al. 2020. Performance evaluation of classifiers for the recognition of online handwrittenGurmukhi characters and numerals: a study. Artificial Intelligence Review 53 (2020), 2075-2097.
13. F. Mushtaq, M. M. Misgar, M. Kumar, and S. S. Khurana. 2021. UrduDeepNet: online handwritten Urdu character recognition using a deepneural network. Neural Computing and Applications (2021), 1-24.
14. P. J. Jino, K. Balakrishnan, and U. Bhattacharya. 2017. Oline Handwritten Malayalam Word Recognition Using a Deep Architecture. InProc. of 7th Int. Conf. on Soft Computing for Problem Solving (SocProS), Vol. 1. 913-925.
15. Sonika Narang, MK Jindal, and Munish Kumar. 2019. Devanagari ancient documents recognition using statistical feature extractiontechniques. Sadhana44, 6 (2019), 1-8.
16. S. R. Narang, M. Kumar, and M. K. Jindal. 2021. DeepNetDevanagari: a deep learning model for Devanagari ancient character recognition.Multimedia Tools and Applications 80, 13 (2021), 20671-20686.
17. H. Singh, R. K. Sharma, V. P. Singh, and M. Kumar. 2021. Recognition of online handwritten Gurmukhi characters using a recurrent neuralnetwork classifier. Soft Computing 25, 8 (2021), 6329-6338.
18. DENG, F., WU, Z., LU, Z., AND BROWN, M. S. BinarizationShop: a user-assisted software suite for convertingold documents to black-and-white. In Proceedings of the 10th annual joint conference on Digital libraries (2010),pp. 255–258.
19. DUA, D., AND GRAFF, C. UCI machine learning repository, 2017.
20. HEDJAM, R., NAFCHI, H. Z., MOGHADDAM, R. F., KALACSKA, M., AND CHERIET, M. Icdar 2015 contest onmultispectral text extraction (ms-tex 2015). In 2015 13th International Conference on Document Analysis andRecognition (ICDAR) (2015), IEEE, pp. 1181–1185.
21. LEWIS, D., AGAM, G., ARGAMON, S., FRIEDER, O., GROSSMAN, D., AND HEARD, J. Building a testcollection for complex document information processing. In Proceedings of the 29th annual international ACMSIGIR conference on Research and development in information retrieval (2006), pp. 665–666.
22. HRADIŠ, M., KOTERA, J., ZEMCIK, P., AND ŠROUBEK, F. Convolutional neural networks for direct textdeblurring. In Proceedings of BMVC (2015), vol. 10.
23. HE, S., AND SCHOMAKER, L. Deepotsu: Document enhancement and binarization using iterative deep learning.Pattern recognition 91 (2019), 379–390.

24. GATOS, B., NTIROGIANNIS, K., AND PRATIKAKIS, I. Icdar 2009 document image binarization contest (dibco2009). In 2009, the 10th International Conference on Document Analysis and Recognition(2009), IEEE, pp. 1375–1382.
25. PRATIKAKIS, I., GATOS, B., AND NTIROGIANNIS, K. H-dibco 2010-handwritten document image binarizationcompetition. In 2010, 12th International Conference on Frontiers in Handwriting Recognition (2010), IEEE,pp. 727–732.
26. GANGEH, M. J., PLATA, M., MOTAHARI, H., AND DUFFY, N. P. End-to-end unsupervised document imageblind denoising. arXiv preprint arXiv:2105.09437 (2021)
27. ZHAO, G., LIU, J., JIANG, J., GUAN, H., AND WEN, J.-R. Skip-connected deep convolutional autoencoder forrestoration of document images. In 2018 24th International Conference on Pattern Recognition (ICPR) (2018),IEEE, pp. 2935–2940
28. GANGEH, M. J., TIYYAGURA, S. R., DASARATHA, S. V., MOTAHARI, H., AND DUFFY, N. P. Documentenhancement system using auto-encoders. In Workshop on Document Intelligence at NeurIPS 2019 (2019)
29. SHARMA, M., VERMA, A., AND VIG, L. Learning to clean: A GAN perspective. In Asian Conference onComputer Vision (2018), Springer, pp. 174–185
30. SOUIBGUI, M. A., AND KESSENTINI, Y. De-GAN: A conditional generative adversarial network for documentenhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence (2020)
31. SOUIBGUI, M. A., KESSENTINI, Y., AND FORNÉS, A. A conditional gan-based approach for distorted camera-captured documents recovery. Pattern Recognition and Artificial Intelligence 1322 (2021), 215
32. HRADIŠ, M., KOTERA, J., ZEMCIK, P., AND ŠROUBEK, F. Convolutional neural networks for direct textdeblurring. In Proceedings of BMVC (2015), vol. 10
33. JEMNI, S. K., SOUIBGUI, M. A., KESSENTINI, Y., AND FORNÉS, A. Enhance to read better: An improvedgenerative adversarial network for handwritten document image enhancement. arXiv preprint arXiv:2105.12710(2021)
34. XU, X., SUN, D., PAN, J., ZHANG, Y., PFISTER, H., AND YANG, M.-H. Learning to super-resolve blurry faceand text images. In Proceedings of the IEEE International Conference on Computer Vision(2017), pp. 251–260
35. CALVO-ZARAGOZA, J., AND GALLEGO, A.-J. A selectional auto-encoder approach for document imagebinarization. Pattern Recognition 86 (2019), 37–47

36. DEY, S., AND JAWANPURIA, P. Light-weight document image cleanup using perceptual loss. arXiv preprintarXiv:2105.09076 (2021)
37. LI, D., WU, Y., AND ZHOU, Y. Sauvolanet: Learning adaptive sauvola network for degraded documentbinarization. arXiv preprint arXiv:2105.05521 (2021)
38. LIN, Y.-H., CHEN, W.-C., AND CHUANG, Y.-Y. Bedsr-net: A deep shadow removal network from a singledocument image. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(2020), pp. 12905–12914
39. TENSMEYER, C., AND MARTINEZ, T. Document image binarization with fully convolutional neural networks.In 2017, the 14th IAPR International Conference on Document Analysis and Recognition (ICDAR) (2017), vol. 1, IEEE,pp. 99–104.
40. MAO, X., SHEN, C., AND YANG, Y.-B. Image restoration using very deep convolutional encoder-decodernetworks with symmetric skip connections. Advances in neural information processing systems 29 (2016),2802–2810
41. WANG, X., YU, F., DUNLAP, L., MA, Y.-A., WANG, R., MIRHOSEINI, A., DARRELL, T., AND GONZALEZ,J. E. Deep mixture of experts via shallow embedding. In Uncertainty in Artificial Intelligence (2020), PMLR,pp. 552–562
42. SAUVOLA, J., AND PIETIKÄINEN, M. Adaptive document image binarization. Pattern recognition 33, 2 (2000),225–236