A SURVEY ON DEEP LEARNING BASED DOCUMENT IMAGE ENHANCEMENT

A PREPRINT

Zahra Anvari
Department of Computer Science and Engineering
University of Texas Arlington
Arlington, TX
zahra.anvari@mavs.uta.edu

Vassilis Athitsos
Department of Computer Science and Engineering
University of Texas Arlington
Arlington, TX
athitsos@uta.edu

January 4, 2022

ABSTRACT

Digitized documents such as scientific articles, tax forms, invoices, contract papers, historic texts are widely used nowadays. These document images could be degraded or damaged due to various reasons including poor lighting conditions, shadow, distortions like noise and blur, aging, ink stain, bleed-through, watermark, stamp, *etc*. Document image enhancement plays a crucial role as a pre-processing step in many automated document analysis and recognition tasks such as character recognition. With recent advances in deep learning, many methods are proposed to enhance the quality of these document images. In this paper, we review deep learning-based methods, datasets, and metrics for six main document image enhancement tasks, including binarization, debluring, denoising, defading, watermark removal, and shadow removal. We summarize the recent works for each task and discuss their features, challenges, and limitations. We introduce multiple document image enhancement tasks that have received little to no attention, including over and under exposure correction, super resolution, and bleed-through removal. We identify several promising research directions and opportunities for future research.

***K*eywords** Document Image Enhancement, Image Enhancement, Document Image Analysis and recognition, Deep Learning

1 Introduction

Digitized documents such as scientific articles, tax forms, invoices, contract papers, personnel records, legal documents, historic texts, *etc.* are ubiquitous and widely used nowadays. These documents can be damaged due to watermark, stamps, aging, ink stains, bleed-through, *etc.*, or can be degraded during the digitization process due to poor lighting conditions, shadow, camera distortion like noise and blur, *etc.* [7, 22, 29, 36, 41, 63].

Degraded document images have low visual quality and legibility. They could contain handwritten or machine printed text, or a mixture of both. In addition, they could contain multiple handwriting styles with different languages. Further complicating matter, the machine used to print the document could have used various technologies with variable quality (*e.g.,* documents printed in low DPI), thus affecting the quality of the image captured. Moreover, old documents could be degraded over time due to different reasons, such as humidity, being washed out, poor storage, low quality medium, *etc.* Therefore, there are many factors that affect the quality and legibility of the digitized document images.

The degraded document images make automated document analysis tasks such as character recognition (OCR) very challenging and such tasks perform poorly on these images. On the other hand, it is impractical and sometimes infeasible to manually enhance such images, especially in large scale, thus it is essential to develop methods that can automatically enhance the visual quality and the legibility of these images and restore the corrupted parts.

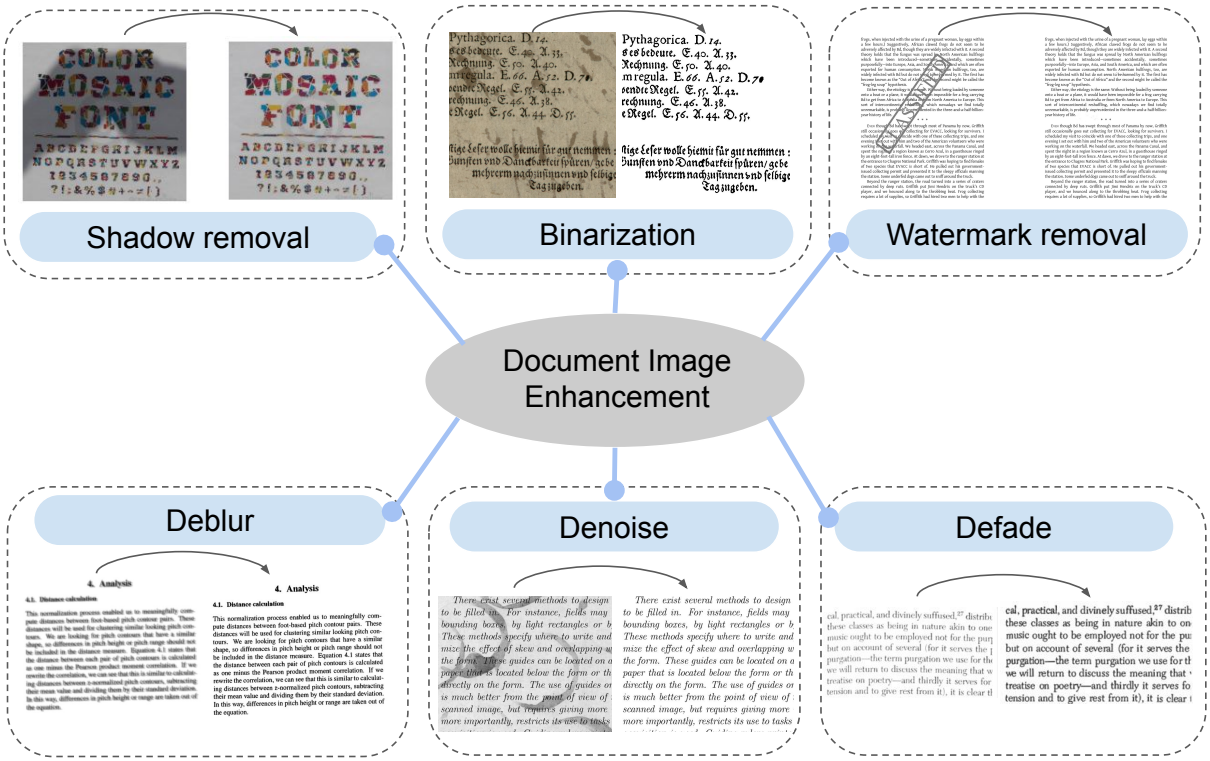


Figure 1: Document image enhancement problems.

Document image enhancement problem consists of several tasks that are studied in the literature. In this survey, we focus on six main tasks that are illustrated in Figure 1, and we explain each task in details in Section 2. Here we summarize these tasks:

- **Binarization:** It aims at separating the background from the foreground (*i.e.,* text) in order to remove noise, ink stain, bleed-through, wrinkles, *etc.* The output of this task is a binary image with two classes: foreground and background.
- **Deblur:** This task aims at removing various blur types, *e.g.,* Gaussian, motion, de-focus, *etc.*, from the document images.
- **Denoise**: Denoising aims at removing various noise types, *e.g.,* salt and pepper, wrinkles, dog-eared, background, and stain, *etc.* from the document images.
- **Defade:** It aims at improving the faded document images. A document could be faded due to againg, overexposure, or being washed out, *etc.*
- **Watermark removal**: Some documents, *e.g.,* financial forms, can contain watermarks, and the text underneath a watermark might not be recognizable. This task aims at removing such watermarks.
- **Shadow removal:** Blocking the source of light while capturing an image (usually by a phone) could leave shadows on the captured document image. This task aims at estimating the show and removing it.

With recent advances in deep learning, deep learning-based approaches have been proposed and applied to different computer vision and image processing tasks, such as object detection [37, 56], semantic segmentation [38], face detection and dataset creation [2, 35, 59], and image enhancement [3, 16, 17], *etc*. It has been shown that such deep learning-based methods achieve promising results and surpass the traditional methods. Similarly, deep learning-based methods for document image enhancement problems have received a great deal of attention over the past few years. The goal of this survey is to review these methods and discuss their features, advantages, disadvantages, challenges, and limitations, and identify opportunities for future research.

To the best of our knowledge, this survey is the first survey of the recent advances in deep learning-based document image enhancement methods. We have several key contributions in this paper:

- We review recent advances, mostly from the past five years, on deep learning-based methods for document image enhancement, to help readers and researchers to better understand this area of research.
- We provide an overview of six main document image enhancement problems, including binarization, deblure, denoise, defade, watermark removal, and shadow removal.
- We review the state-of-the-art methods, and discuss their features, advantages, and disadvantages to help researchers and investigators to select suitable methods for their need.
- We introduce several important document image enhancement tasks that have received little to no attention, such as bleed through removal.
- We identify several open problems and promising research directions and opportunities for future research.

2 Document Image Enhancement Tasks

In this section, we describe six main document image enhancement tasks, including binarization, debluring, denoising, defading, watermark removal, and shadow removal. Figure 2 shows some image examples for each of these tasks.

2.1 Binarization

Document image binarization refers to the process of segmenting a gray scale or color image to a black-and-white or binary image with only text and background. During this process any existing degradations such as bleed-through, noise, stamp, ink stains, faded characters, artifacts, etc. are removed. Formally, it seeks a decision function $f_{binarize}()$ for a document image D_{orig} of width W and height H, such that the resulting image $D_{binarized}$ of the same size only contains binary values while the overall document legibility is at least maintained if not enhanced.

$$D_{binarized} = f_{binarize}(D_{orig}) \tag{1}$$

Figure 2a shows an example of an image along with its binarized one.

2.2 Debluring

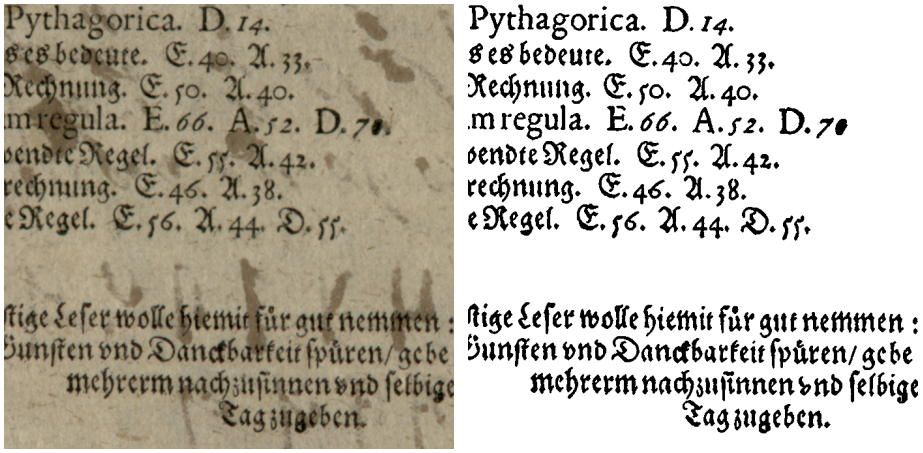
Nowadays, smartphones are widely used to digitize documents. This might propose various issues. The most prevalent one is the blur that might be introduced during capturing process. For instance, movement of the document, camera being out of focus, and camera shakes can add blur to the captured image. Figure 2b shows an example of a blurry document image along with its corresponding clean one.

The goal of deblurring methods is to recover the clean or deblurred version of the blurry document image. These methods could be prior-based or learning based. The former ones attempt to estimate the blur kernel and the corresponding parameters to detect blur and use these parameters to remove it, thus recover the clean images. The learning-based methods which are also called data-driven methods are widely used in the past decade. These methods take advantage of the deep neural networks and large amount of data to propose a deblurring model that can recover the clean image without requiring any priors.

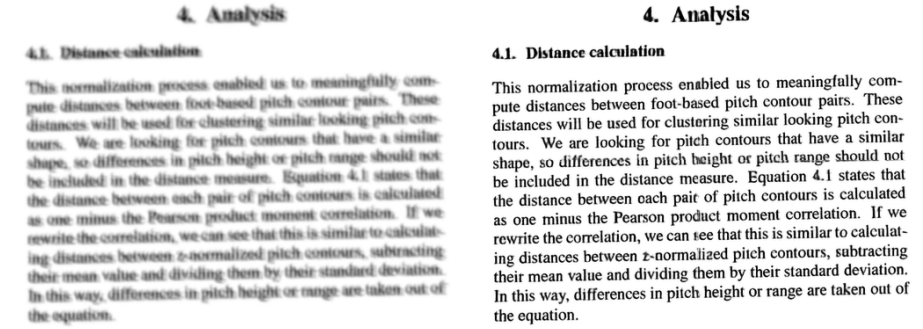
Document image deblurring is an ill-posed problem and it is a more challenging problem compared to natural/non-document image deblurring. One of the main reasons is that the performance of the OCR engines directly depend on the quality of the document images that are input to them. If the legibility and the quality of these document images were low, the performance of the OCR output will be affected accordingly. Therefore, the enhanced document images not only need to be visually improved they also need to become more legible.

2.3 Defading

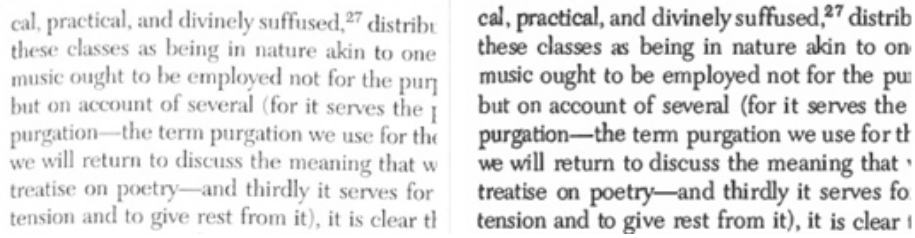
Defading is the process of recovering documents' text that have become faded/faint. Documents' content can be faded due to different factors. For instance, the ink can wear off over time, which is more prevalent in old documents. Sun-light or overexposure while digitizing the document can also make the document content lightened and hard to read. In addition, the handwriting or the printed text can be faint in the first place and deteriorate over time. This type of degradation poses issues such as low visual quality, poor legibility, and poor OCR performance. Defading methods mainly attempt to increase the visibility and recover a more legible version of the document image. Figure 2c shows an example of a defaded document image and its corresponding ground truth.



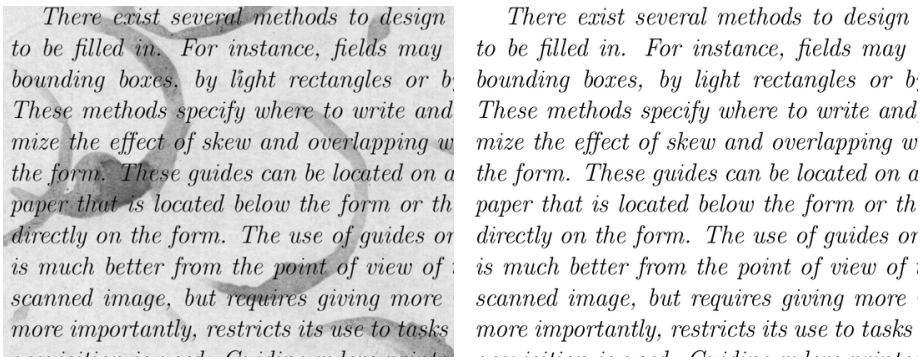
(a) Binarization task [55].



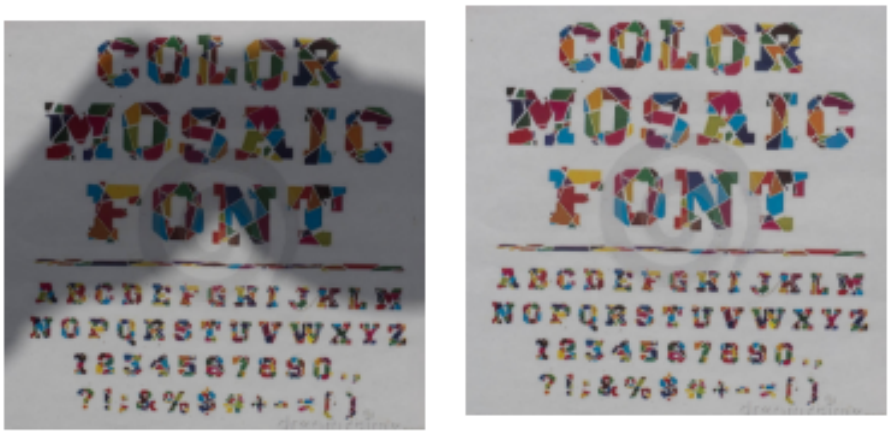
(b) Deblur task. [23]



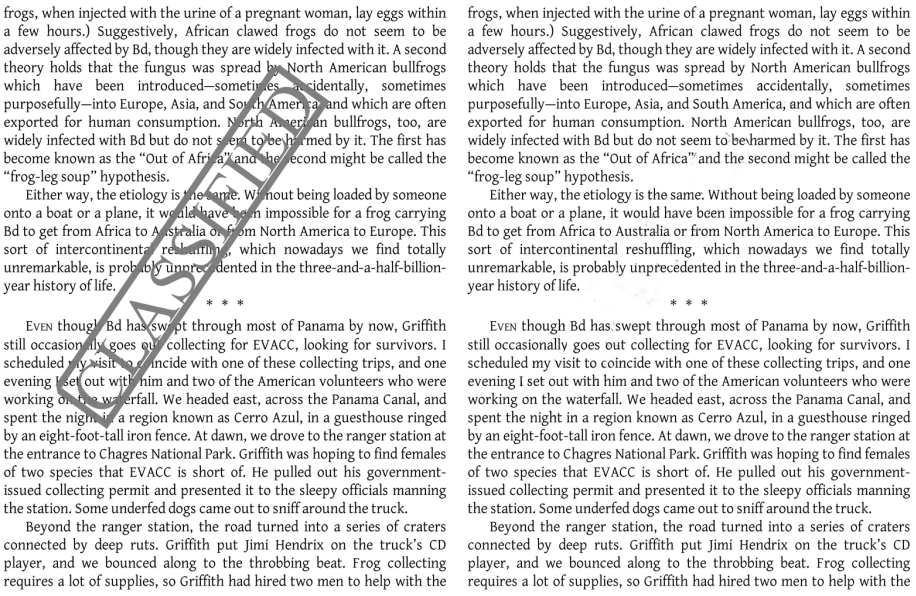
(c) Defade task.



(d) Denoise task. [32]



(e) Shadow removal task. [36]



frogs, when injected with the urine of a pregnant woman, lay eggs within a few hours.) Suggestively, African clawed frogs do not seem to be adversely affected by Bd, though they are widely infected with it. A second theory holds that the fungus was spread by North American bullfrogs which have been introduced—sometimes accidentally, sometimes purposefully—into Europe, Asia, and South America, and which are often exported for human consumption. North American bullfrogs, too, are widely infected with Bd but do not seem to be harmed by it. The first has become known as the "Out of Africa" and the second might be called the "frog-leg soup" hypothesis.

Either way, the etiology is the same. Without being loaded by someone onto a boat or a plane, it would have been impossible for a frog carrying Bd to get from Africa to Australia or from North America to Europe. This sort of intercontinental reshuffling, which nowadays we find totally unremarkable, is probably unprecedented in the three-and-a-half-billion-year history of life.

* * *

Even though Bd has swept through most of Panama by now, Griffith still occasionally goes out collecting for EVACC, looking for survivors. I scheduled my visit to coincide with one of these collecting trips, and one evening I set out with him and two of the American volunteers who were working on the waterfall. We headed east, across the Panama Canal, and spent the night in a region known as Cerro Azul, in a guesthouse ringed by an eight-foot-tall iron fence. At dawn, we drove to the ranger station at the entrance to Chagres National Park. Griffith was hoping to find females of two species that EVACC is short of. He pulled out his government-issued collecting permit and presented it to the sleepy officials manning the station. Some underfed dogs came out to sniff around the truck.

Beyond the ranger station, the road turned into a series of craters connected by deep ruts. Griffith put Jimi Hendrix on the truck's CD player, and we bounced along to the throbbing beat. Frog collecting requires a lot of supplies, so Griffith had hired two men to help with the

(f) Watermark removal task [61]

Figure 2: Sample images for different document image enhancement tasks. The image on the left is the input, and the right image is the output of each task.

2.4 Denoising

Some documents may contain artifacts such as salt and pepper noise, stamps, annotations, ink or coffee stains, wrinkles, *etc.* The image recovery is even harder when certain types of these artifacts cover the text specially in cases where the artifacts color is similar to or darker than the document text color. To improve the visual quality of these document images alongside the legibility, approaches that recover the clean version of the degraded documents are proposed. The methods that attempt to remove these artifacts include document image denoising, cleanup and binarization methods. Figure 2d illustrates an example of a noisy document image and its ground truth.

2.5 Shadow Removal

Documents can be digitized using scanners or mobile phone cameras. In the past, scanners were commonly used for digitizing documents with high quality, but with the prevalence of mobile phones more people tend to use their phones cameras in place of scanners to capture digital copies of their documents.

Dataset	Task	No. of images	Resolution(Pixels)	Real vs. synthetic
Bishop Bickley diary [9]	Binarization	7	1050 x 1350	Real
NoisyOffice [12]	Denoising	288	Variable	Real/Synthetic
S-MS [21]	Multiple	240	1001 x 330	Synthetic
Tobacco 800 [32]	Denoising	1290	(1200x1600) - (2500x3200)	Real
DIBCO'17	Binarization	10	(1050x608) - (2092x951)	Real
H-DIBCO'17	Binarization	10	(351x292) - (2439x1229)	Real
SmartDoc-QA [44]	Deblurring	4260	-	Real
Blurry document images [23]	Deblurring	3M train/35K validation	300 x 300	Synthetic

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

The document images captured using mobile phones are vulnerable to shadows mainly because the light sources are often blocked by the camera or even the person's hand. Furthermore, even in the absence of objects that could be a source of occlusion, the lighting is often uneven when the document image is being captured in the real life. Therefore, document images digitized by mobile phone cameras in particular can suffer from shadows blocking a portion or all of the document and also uneven lighting and shading. These result in poor visual quality and legibility. Shadow removal methods focus on estimating the shadow casted on the document image and attempt to remove that in order to recover a clean, evenly lit document image which is more legible than the shadowed version. Figure 2e presents a sample of a document image with shadow and its ground truth.

2.6 Watermark Removal

Some documents, *e.g.,* financial forms, may contain one or multiple watermarks which occlude the document texts or makes it hard to read. Similar to denoising, the document image recovery is even harder in cases where the watermark color is the same or darker than the document text color or the watermark is thick and dense. Hence, we need approaches that recover the clean version of the degraded documents. Watermark removal methods focus on removing watermarks in order to increase the visual quality and legibility of the document images. Figure 2f shows an image sample along with its ground truth for this task.

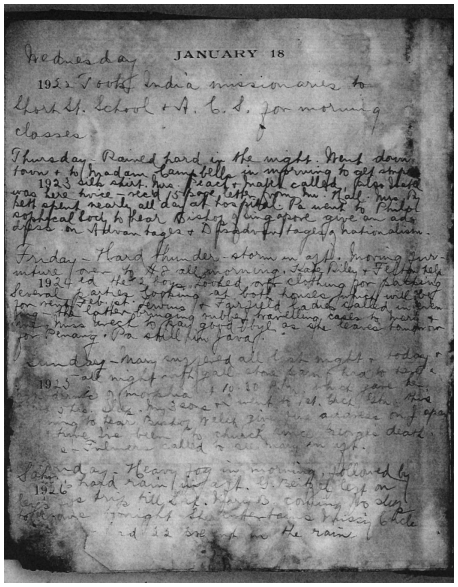
3 Datasets

In this section, we describe datasets that are used in the literature for different document image enhancement tasks. Table 1 provides the specifications of these datasets and we describe them in more details in below. In addition, Figure 3 shows image samples from these datasets.

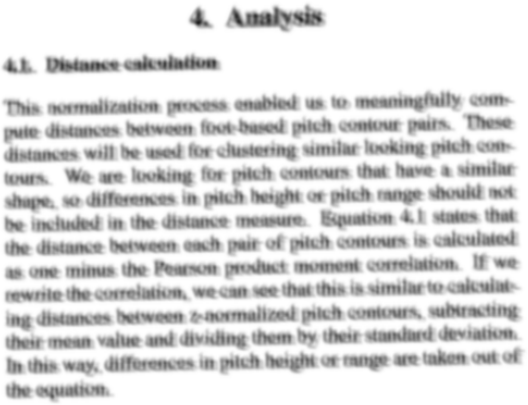
Bickley diary [9]: The images of Bickley diary dataset are taken from a photocopy of a diary that is written about 100 years ago. These images suffer from different kinds of degradation, such as water stains, ink bleed-through, and significant foreground text intensity. This dataset contains 7 document images/pages along with the binarized/clean ground truth images.

NoisyOffice [12]: 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. The images in this dataset contain various styles of text, to which synthetic noise has been added to simulate real-world, messy artifacts.

S-MS (Synchromedia MultiSpectral Ancient document) [21]: Multi-spectral imaging (MSI) represents an innovative and non-destructive technique for the analysis of materials such as ancient documents. They collected a multispectral image database 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. Original documents were borrowed from Quebec's national library and have been imaged using a CROMA CX MSI camera. Through this process, they produced 8 images for each document resulting in total of 240 images of real documents.



(a) Sample image from Bickley Diary Dataset [9]

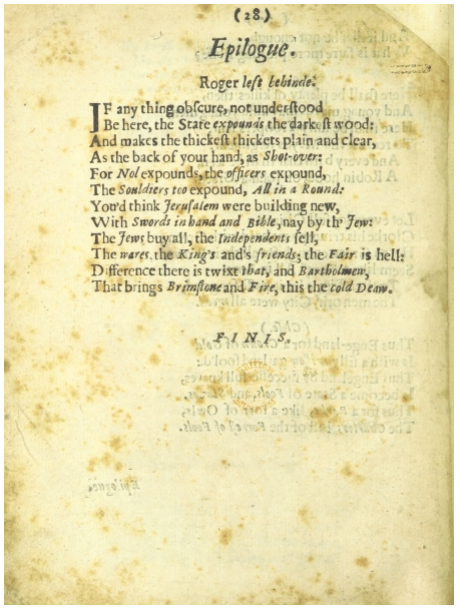


4. Analysis

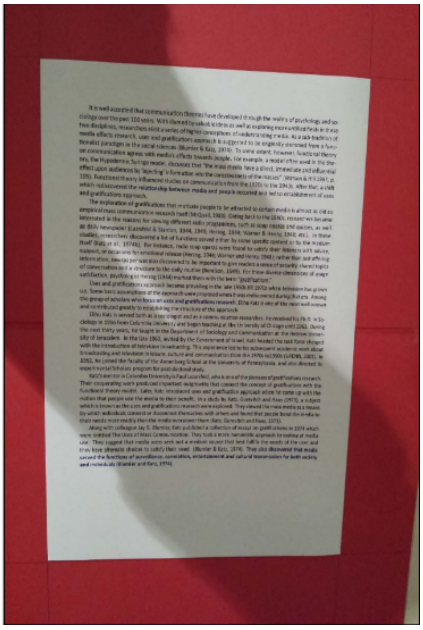
4.1. Distance calculation

This normalization process enabled us to meaningfully compute distances between foot-based pitch contour pairs. These distances will be used for clustering similar looking pitch contours. We are looking for pitch contours that have a similar shape, so differences in pitch height or pitch range should not be included in the distance measure. Equation 4.1 states that the distance between each pair of pitch contours is calculated as one minus the Pearson product moment correlation. If we rewrite the correlation, we can see that this is similar to calculating distances between z-normalized pitch contours, subtracting their mean value and dividing them by their standard deviation. In this way, differences in pitch height or range are taken out of the equation.

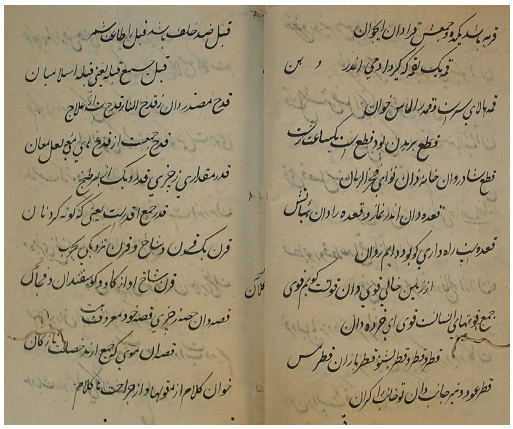
(b) Sample image the dataset introduced in [23]



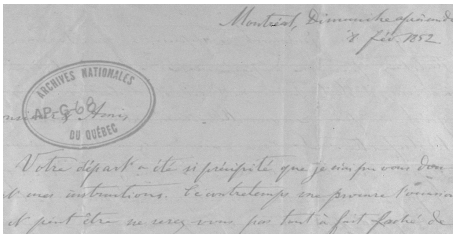
(c) Sample image from DIBCO Dataset [55]



(d) Sample image from SmartDoc-QA Dataset [44]

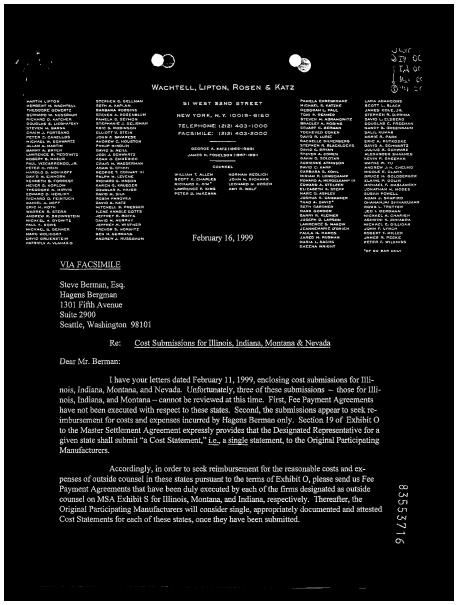


(e) Sample image from PHIDB dataset [43]

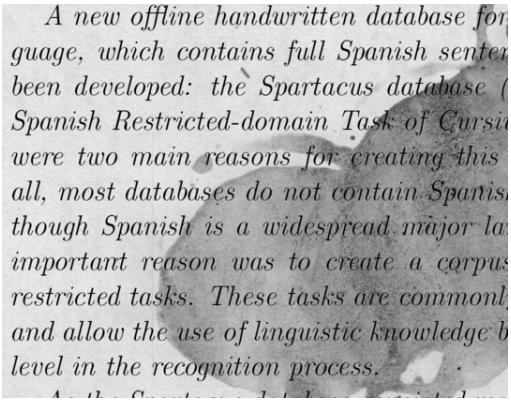


(f) Sample image from S-MS Dataset [21]

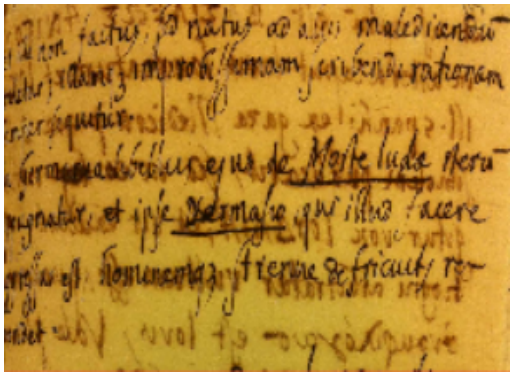
Figure 3: Sample images from datasets for document image enhancements tasks.



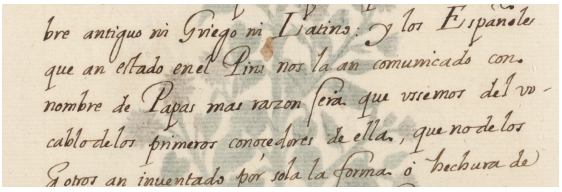
(g) Sample image from Tobacco Dataset [32]



(h) Sample image from Noisy Office Dataset [32]



(i) Sample image from MCS dataset [20]



(j) Sample image from H-DIBCO Dataset [51]

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

Tobacco 800 [32]: This is a publicly available subset of 42 million pages of documents that are scanned with various equipment. It contains real-world documents with different types of noise and artifact, such as stamps, handwritten texts, and ruling lines, on the signatures. Resolutions of documents in Tobacco800 vary significantly from 150 to 300 DPI and the resolution of the document images vary from 1200x1600 to 2500x3200 pixels.

DIBCO and H-DIBCO: These datasets were introduced for the Document Image Binarization Contest since 2009. There are DIBCO 2009 [15], H-DIBCO 2010 [49], DIBCO 2011 [50], H-DIBCO 2012 [51], DIBCO 2013 [52], H-DIBCO 2014 [46], H-DIBCO 2016 [54], DIBCO 2017 [55], H-DIBCO 2014 [46], H-DIBCO 2018 [53]. DIBCO datasets contain both printed and handwritten document images mainly for the binarization task.

SmartDoc-QA [44]: This is a dataset for quality assessment of smartphone captured document images containing both single and multiple distortions. This dataset is created using smartphone's camera captured document images, under varying capture conditions such as light, shadow, different types of blur and perspective angles. SmartDoc-QA is categorized in three subsets of documents: contemporary documents, old administrative documents and shop's receipts.

Blurry document images (BMVC) [23]: 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.

Monk Cuper Set (MSC) [20]: This dataset contains 25 pages sampled from real historical documents which are collected from the Cuper book collection of the Monk system [68]. MSC documents suffer from heavy bleed-through degradations and textural background.

Persian heritage image binarization dataset (PHIDB) [43]: The PHIBD 2012 dataset contains 15 historical document images with their corresponding ground truth binary images. The historical images in this dataset suffer from various types of degradation. In particular two types of foreground text degradation are nebulous, and weak strokes/sub-strokes and the background degradation types are global bleed-through, local bleed-through, unwanted lines/patterns, and alien ink.

4 Metrics

In this section, we describe the evaluations metrics that are used in the literature for different document image enhancement tasks.

- **Peak signal-to-noise ratio (PSNR)**: PSNR is a referenced-based metric. It provides a pixel-wise evaluation and is capable of indicating the effectiveness of document enhancement methods in terms of visual quality. PSNR measures the ratio between the maximum possible value of a signal and the power of distorting noise that affects the quality. In other words, it measures the closeness of two images. The higher the value of PSNR, the higher the similarity of the two images. MAX is the maximum possible pixel value of the image. When the pixels are represented using 8 bits per sample, MAX is 255. Given two MxN images, this metric would be formulated as follows:

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

  where

$$MSE = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}(I(x,y) - I^{'}(x,y))^2}{MN} \tag{3}$$

- **Structural Similarity Index (SSIM) [72]**: SSIM is a reference-based metric designed to measure the structural similarity between two images and quantifies image quality degradation. SSIM computation requires two images from the same image, a reference image and a processed image. It actually measures the perceptual difference between two similar images. This metric extracts three key features from an image: luminance, contrast, and structure. The comparison between the two images is performed on the basis of these three features.
- **Character Error Rate (CER)**: Character Error Rate is computed based on the Levenshtein distance. It is the minimum number of character-level operations required to transform the ground truth or reference text into the OCR output text. CER is formulated as follows:

$$CER = \frac{S + D + I}{N} \tag{4}$$

  where S is the number of Substitutions, D is the number of Deletions, I is the number of Insertions, and N is the number of characters in reference or ground truth text.

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

$$CER_{normalized} = \frac{S + D + I}{S + D + I + C} \tag{5}$$

- **Word Error Rate (WER)**: Word Error Rate can be more used for evaluating the OCR performance on paragraphs and sentences. WER is formulated in below:

$$WER = \frac{S_w + D_w + I_w}{N} \tag{6}$$

  WER is computed similar to CER, but WER operates at word level. It represents the number of word substitutions, deletions, or insertions needed to transform one sentence into another.
- **F-measure [52]**: The F-measure score is the harmonic mean of the precision and recall. Precision is the positive predictive value, and recall aka sensitivity is used in binary classification. F-measure is formulated as follows:

$$FM = \frac{2 \times Recall \times Precision}{Recall + Precision} \tag{7}$$

where

$$Recall = \frac{TP}{TP + FN} \tag{8}$$

$$Precision = \frac{TP}{TP + FP} \tag{9}$$

TP, FP, FN denote the True Positive, False Positive and False Negative values, respectively.

- **Pseudo-FMeasure (F_{ps}) [52]**: F_{ps} is introduced in [45] and it utilizes pseudo-recall Rps and pseudo-precision P_{ps}. It follows the same formula as F-Measure explained above and is particularly used for the binarization task.

  In the case of pseudo-recall, the weights of the ground truth(GT) foreground are normalized according to the local stroke width. Generally, those weights are between [0,1]. In the case of pseudo-precision, the weights are constrained within an area that expands to the GT background taking into account the stroke width of the nearest GT component. Inside this area, the weights are greater than one (generally between (1,2]) while outside this area they are equal to one.

- **Distance Reciprocal Distortion Metric (DRD) [52]**: DRD metric is used to measure the visual distortion in binary document images [39]. It correlates with the human visual perception and it measures the distortion for all pixels as follows:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{10}$$

  where NUBN is the number of the non-uniform 8x8 blocks in the GT image, and DRD_k is the distortion of the kth flipped pixel that is calculated using a 5x5 normalized weight matrix W_{Nm} as defined in [39]. DRD_k equals to the weighted sum of the pixels in the 5x5 block of the GT that differ from the centered kth flipped pixel at (x, y) in the binarization result image (equation 11).

$$DRD_k = \sum_{i=-2}^{2} \sum_{j=-2}^{2} |GT_k(i, j) - B_k(x, y)| \times W_{Nm}(i, j) \tag{11}$$

5 Document Image Enhancement Methods

In this section, we describe the main deep learning based methods for document image enhancement and discuss their features, challenges, and limitations. Most of these works focused on multiple tasks, therefore in this section we discuss the document enhancement methods chronologically. Table 3 summarizes the advantages, disadvantages, and results of these methods. Below, we describe these methods in more details.

The method introduced in [23] is proposed for document image deblurring problem. The authors proposed a small and computationally efficient convolutional neural network model to deblur images without assuming any priors. In particular the authors focused on a combination of realistic de-focus blur and camera shake blur. They demonstrated that the proposed network significantly outperform existing blind deconvolution methods both in terms of image quality, PSNR, and OCR accuracy, CER. The proposed model can also be used on mobile devices as well.

In another document image deblurring work [73], the authors proposed an algorithm to directly restore a high-resolution de-blurred image from a blurry low-resolution input. Other deblurring methods such as Hradis et al. [23] cannot be easily extended for joint super-resolution and deblurring tasks. This work focuses on blurry face and blurry document images distributions and a multi-class GAN model was developed to learn a category-specific prior and process multi-class image restoration tasks, using a single generator network. The authors employed a deep CNN architecture proposed by Hradis et al. [23] in an adversarial setting. Unlike Hradis et al., in this work the generator network contains upsampling layers, which are fractionally-strided convolutional layers aka deconvolution layers. The generator first upsamples low-resolution blurry images, and then performs convolutions to generate clear images thus the output would be both super-resolved and deblurred. Since their model has a discriminator network in addition to the generator network, it is more complex and has more parameters compared to model proposed in [23].

The visual quality of the generated images were evaluated in terms of PSNR and SSIM but the deblurred document images were not evaluated in terms of OCR performance and no Character Error Rate or Word Error Rate which are OCR performance evaluation metrics are reported. In terms of PSNR/SSIM, this work performs favorably against previous work on both synthetic and real-world datasets.

Methods	Document Image Enhancement Tasks						Document Type	
	Binarization	Deblur	Denoise	Defade	Watermark Removal	Shadow Removal	Handwritten	Printed
Gangeh et al. [13]	-	✓	✓	✓	✓	-	-	-
Zhao et al. [74]	-	✓	✓	-	-	-	-	-
Sharma et al. [60]	-	✓	-	✓	✓	-	-	✓
Lin et al. [36]	-	-	-	-	-	✓	✓	-
Souibgui et al. [61]	-	✓	-	-	✓	-	✓	✓
Gangeh et al. [14]	-	✓	-	-	✓	-	-	✓
Hradiš et al. [23]	-	✓	-	-	-	-	-	-
Jemni et al. [25]	✓	-	-	-	-	-	✓	-
Xu et al. [73]	✓	-	-	-	-	-	-	✓
Souibgui et al. [62]	-	✓	-	-	-	✓	-	✓
Calvo-Zaragoza et al. [6]	✓	-	-	-	-	-	✓	✓
Dey et al. [10]	✓	-	✓	-	-	-	-	✓
Li et al. [33]	✓	-	-	-	-	-	✓	✓

(a) Tasks and document types handled by the of main methods reviewed in this paper

Methods	GAN	CNN	Paired vs. unpaired supervision
Gangeh et al. [13]	✓	-	Unpaired
Zhao et al. [74]	-	✓	Paired
Sharma et al. [60]	✓	-	Unpaired
Lin et al. [36]	✓	-	Paired
Souibgui et al. [61]	✓	-	Paired
Gangeh et al. [14]	-	✓	Paired
Hradiš et al. [23]	-	✓	Paired
Jemni et al. [25]	✓	-	Paired
Xu et al. [73]	✓	-	Paired
Souibgui et al. [62]	✓	-	Paired
Calvo-Zaragoza et al. [6]	-	✓	Paired
Dey et al. [10]	-	✓	Paired
Li et al. [33]	-	✓	Paired

(b) Methodologies used in the reviewed methods.

Table 2: Description of main methods reviewed in this paper.

Methods	Advantages	Disadvantages	Results
Gangeh et al. [13]	- Handles multiple noises including salt and pepper noise, faded, blurred, and watermarked documents in an end-to-end manner. - It does not rely on paired document images.	- Computationally complex.	- Method has best results in terms of PSNR and OCR as compared to previous three methods.
Zhao et al. [74]	- Method is fast and easy to implement.	- Inadequate qualitative and qualitative results.	- Marginal PSNR improvement.
Sharma et al. [60]	- Adaptable for both paired and unpaired supervision scenarios.	-	- Marginal improvement in terms of PSNR.
Lin et al. [36]	- First deep learning-based approach for shadow removal. - It works on both gray-scale and RGB images.	- Computationally complex. - It does not work well on images with complex background and layouts. - It works well on partially shadowed documents only.	- It achieves the best results in terms of PSNR/SSIM compared to four previous work when evaluated on five different datasets. - It also generalizes relatively well on real-world images.
Souibgui et al. [61]	- 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 less practical.	**Binarization:** Achieves best results in terms of PSNR, $F_{measure}$, F_{ps} and DRD compared to top five competitors. **Watermark:** Achieves best results in terms of PSNR/SSIM compared to three previous work. **Deblur:** Achieves best results in terms of PSNR compared to two previous work.
Gangeh et al. [14]	- Works on both gray-scale and RGB watermarks. - Works on blurry images with various intensity.	- Inadequate quantitative evaluation and comparison with previous work.	- Effectively removes watermark and blur. - Improved OCR on a small test set of nine images.
Hradiš et al. [23]	- Small and computationally efficient network. - Can be used on mobile devices.	- Adds ringing artifacts in some situations. - Does not work well on uncommon words when the image is severely blurred.	- Outperforms other methods in terms of PSNR and Character Error Rate compared to previous four work.
Xu et al. [73]	- Computationally efficient network. - It deblurs and super-resolves simultaneously.	- Does not generalize well for generic images. - OCR performance evaluation is ignored and only visual quality of the documents are evaluated.	- Performs favorably against previous work on both synthetic and real-world datasets.
Souibgui et al. [62]	- It handles multiple camera distortions. - It incorporates a text recognizer for generating more legible images.	- Model only processed and trained on single lines and can not handle full pages.	- Achieves best results in terms of Character Error Rate and second best in terms of PSNR/SSIM compared to previous three work.

Table 3: Comparison of document image enhancement methods.

One limitation of this work is that since the model is trained on multi-class images, it is essentially designed to approximate the mixture distribution of these two classes of images and when this mixture distribution becomes too complex, it is difficult to learn a unified model to cover the diversity of all image classes. Therefore, this method is less effective for generic images.

Authors in [66] focused on the degraded historical manuscript images binarization, and formulated binarization task as a pixel classification learning task. They developed a Fully Convolutional Network (FCN) architecture that operates at multiple image scales, including full resolution. The authors claimed that the proposed binarization technique can also be applied to different domains such as Palm Leaf Manuscripts with good performance.

Zhao et. al. [74] investigated the denoising and deblurring problems and proposed a method for document image restoration 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 is defined to connect the input to the output directly. These connections assist the network to learn the residual content between the noisy and clean images instead of learning an ordinary transformation function. The proposed network was inspired by [23] which is a 15-layer CNN. Compared to method in [23], the authors added batch normalization [24] and skip-connections [19] to accelerate the model convergence of the model and boost the performance.

In [60], the authors cast the image restoration problem as an image-to-image translation task i.e, translating a document from noisy domain (*i.e.,* background noise, blurred, faded, watermarked) to a target clean document using a GAN approach. To do so, they employed 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 and defading problems by inserting logos as watermarks and applying fading techniques on Google News dataset [67] of documents.

Authors in [14] proposed an end-to-end document enhancement pipeline which takes in blurry and watermarked document images and produces clean documents. They trained an auto-encoder model that works on different noise levels of documents. They adopted the neural network architecture described in [40] called REDNET and designed a REDNET with 15 convolutional layers and 15 deconvolutional layers, including 8 symmetric skip connections between alternate convolutional layers and the mirrored deconvolutional layers. The advantage of this method compared to fully convolutional network is that pooling and un-pooling, which tend to eliminate image details, is avoided for low-level image tasks such as image restoration. This results in higher resolution outputs. The key differences of this work from [74] is the use of larger dataset and training a blind model.

In [67] authors developed convolutional auto-encoders to learn an end-to-end map from an input image to its selectional output, in which the activations indicate the likelihood of pixels to be either foreground or background. Once trained, this model can be applied to documents to be binarized and then a global threshold will be applied. This approach has proven to outperform existing binarization strategies in a number of document types.

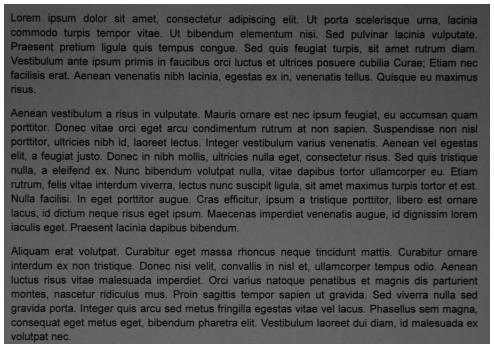
In DE-GAN [61], the authors proposed an end-to-end framework called Document Enhancement Generative Adversarial Networks. This network is based on conditional GANs and cGANs, a network to restore severely degraded document images. The tasks that are studied in this paper are document clean up, binarization, deblurring and watermark removal. Due to unavailability of a dataset for the watermark removal task, the authors synthetically created a watermark dataset including the watermarked images and their clean ground truth.

Authors in [36] 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 estimation module for extracting the global background color of the document. During the process of estimating the background color, 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 and the 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 with a color gradient and another case is when the document is entirely shadowed, or multiple shadows were formed by multiple light sources.

In another work [62] the authors focused on documents that are digitized using smart phone's cameras. They stated that these types of digitized documents are highly vulnerable to capturing various distortions including but not limited to perspective angle, shadow, blur, warping, etc. The authors proposed a conditional generative adversarial network that maps the distorted images from its domain into a readable domain. This model integrates a recognizer in the discriminator part for better distinguishing the generated document images.

In another study [13], 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 was proposed. In particular they proposed a

(a) Over-exposure problem. Image obtained from [4].



(b) Under-exposure problem. Image obtained from [42].

Figure 4: Open problems: over-exposure and under-exposure correction.

unified architecture by integrating deep mixture of experts [70] with a cycle-consistent GAN as the base network for document image blind denoising problem.

In [10], authors target 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 CNN architecture, incorporated with perceptual loss. They proved that in terms of the number of parameters and product-sum operations, their models are 65-1030 and 3-27 times, respectively, smaller than existing SOTA document enhancement models.

In another work [25], authors focused on enhancing handwritten documents and proposed an end-to-end GAN-based architecture to recover the degraded documents. Unlike most document binarization methods, which only attempt to improve the visual quality of the degraded document, the proposed architecture integrates a handwritten text recognizer that promotes the generated document image to be also more legible. This approach is the first work to use the text information while binarizing handwritten documents. They performed experiments on degraded Arabic and Latin handwritten documents and showed that their model improves both the visual quality and the legibility of the degraded document images.

In [33], authors proposed a document binarization method called SauvolaNet. They investigated the classic Sauvola [58] document binarization method from the deep learning perspective and proposed a multi-window Sauvola model. They also introduced an attention mechanism to automatically estimate the required Sauvola window sizes 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 Sauolva Threshold. 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 that is in charge of estimating the preferred window sizes for each pixel. The other module, Adaptive Sauolva Threshold, combines the outputs from the other two modules and predicts the final adaptive threshold for each pixel. The SauvolaNet model significantly reduces the number of required network parameters and achieves SOTA performance for document binarization task.

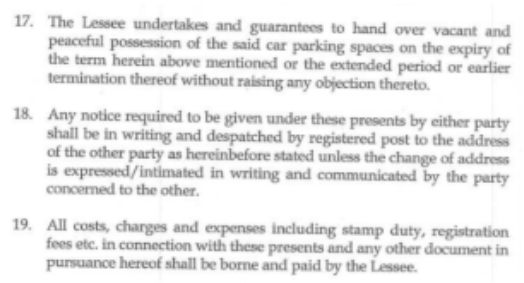
6 Open Problems and Future Directions

In this section, we present open problems in this area and provide several directions for the future work. Document image enhancement tasks are far from solved and even some tasks are either not studied or studied in a very limited fashion. We discuss these problems and future work below.

6.1 Overexposure and underexposure correction tasks

Overexposure problem occurs when too much light is captured while digitizing the document, mostly when the capturing device is a mobile phone and camera flash adds too much reflection or glare to the image (Figure 4a). This problem has received limited attention even in the image and photo enhancement domain [1, 5], and to the best of our knowledge no study has tried to address this problem for the document images. To address this issue with a deep learning based approach, training and testing datasets are required to be collected, as no public datasets are available that can be leveraged for this problem.

On the other hand, underexposure happens when the lighting condition is poor while digitizing the document and as a result the captured image becomes dark (Figure 4b). This problem is different from shadow removal, as shadowed

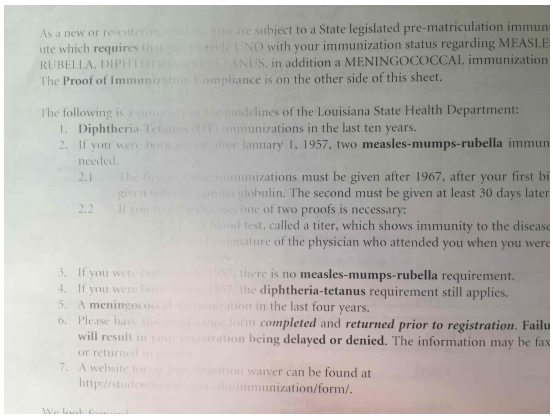


17. The Lessee undertakes and guarantees to hand over vacant and peaceful possession of the said car parking spaces on the expiry of the term herein above mentioned or the extended period or earlier termination thereof without raising any objection thereto.

18. Any notice required to be given under these presents by either party shall be in writing and despatched by registered post to the address of the other party as hereinbefore stated unless the change of address is expressed/intimated in writing and communicated by the party concerned to the other.

19. All costs, charges and expenses including stamp duty, registration fees etc. in connection with these presents and any other document in pursuance hereof shall be borne and paid by the Lessee.

(a) Sample slightly faded image studied in the literature [13].



(b) Sample real-world image with severe and non-homogeneous fade. These types of fade is not studied in the literature.

Figure 5: Open problem in defading task: Severely and/or non-homogeneously faded images.

document images can be partly/non-uniformly dark [36]. While low-light image enhancement problem received a lot of attention for photos [18, 26, 57, 69], it has not received much attention in document image enhancement [34]. One possible future work could be to evaluate the practicality of these methods over document images. Similar to overexposure correction task, developing deep learning based methods for this problem needs training/testing datasets, but such datasets are unavailable.

6.2 Defading task

Fading could occur due to exposure to light, againg, being washed out, *etc*. This task is yet another ill-posed and under-studied task. Current work [13] makes two assumptions that may not be practical. They assume that the documents are uniformly faded, and the documents are very lightly faded (Fig. 5a), while in real-world scenarios the documents could be severely and/or non-homogeneously faded, *e.g.,* aged or washed out documents (Fig. 5b).

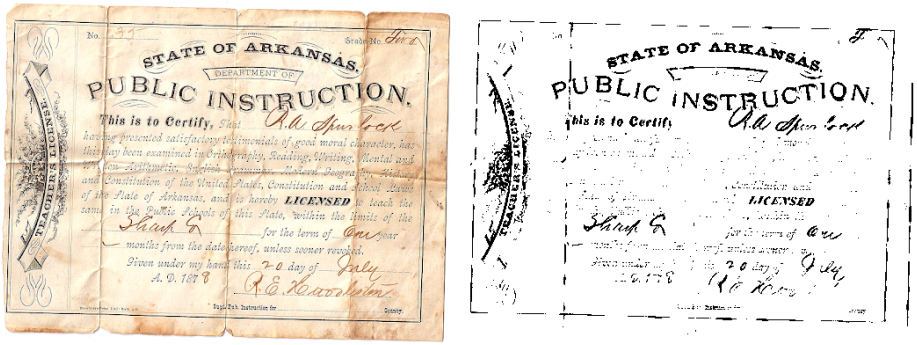
Heavily and/or non-homogeneously faded documents are hard to read and very challenging for OCR and could considerably affect the performance of OCR, while lightly faded documents are usually still legible and recognizable to OCR. Therefore, to address these challenges we need to develop solutions that would take into account both severely and non-homogeneously faded documents. In addition, to train deep learning models (for both lightly and severely faded documents) training datasets are required, but similar to previous task discussed above no such datasets are publicly available.

6.3 Super-resolution task

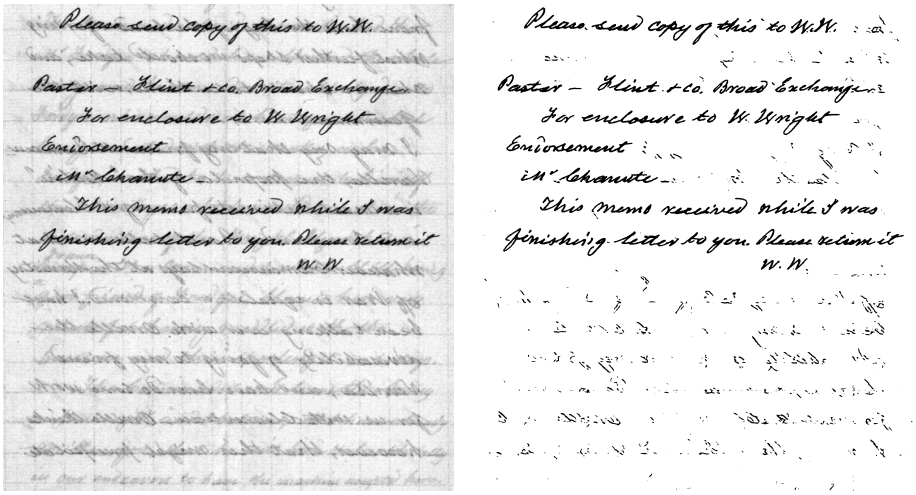
Low-resolution documents are often hard to read and also very challenging to character recognition methods. Super-resolving low resolution document images can enhance the visual quality, readability of the text, and more importantly improve the OCR accuracy. Document image super-resolving is an ill-posed and challenging problem, especially when there are artifacts and noises present in the documents. Developing a model that super-resolves the document images in particular low-quality document images is even harder and more challenging.

One way to tackle this issue is to use Bicubic interpolation but such basic methods can introduce noise or exacerbate the noise/artifacts that the document in particular low-quality ones have. To increase the resolution of the document images and recover as much details as possible we need super resolution methods. Through super-resolving these document images, characters become more legible and it could boost the OCR performance as well.

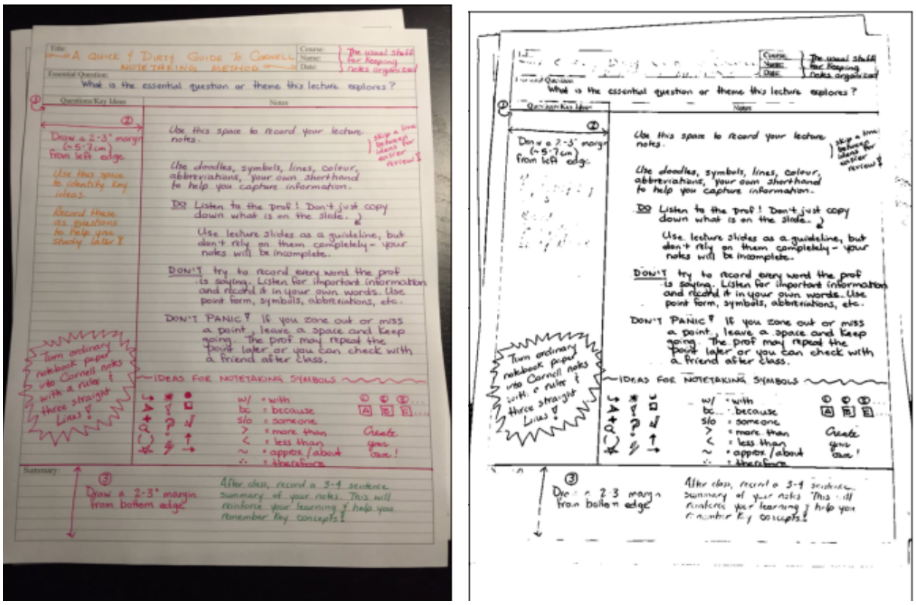
While image/photo super-resolution problem has received a great deal of attention [8, 11, 27, 28, 30, 31, 40, 64, 65, 71], this task has received little attention for the document images [47, 48]. As a future work, we need to develop effective super-resolution methods specifically designed for documents image with low-quality in order to improve the legibility and OCR performance.



(a) Low contrast problem. Low contrast text is not recovered.



(b) Bleed-through problem. Bleed through text is not effectively removed.



(c) RGB document with various ink intensities. Text in lighter color i.e. orange, is not effectively recovered.

Figure 6: Open problems in binarization task. Left side images are the original images, and the right side images are the binarized images using a recent state-of-the-art method.

6.4 Binarization task

While the binarization task has received a great deal of attention, there are still multiple scenarios that current binarization methods do not perform well on them. Specifically, when the image has low contrast or when ghosting and bleed through are present in the document, or when the image is RGB with various ink color and intensities. These scenarios are challenging for binarization methods to handle.

Ghosting in documents occurs when the ink or text from the other side of the page can be seen, but ink does not completely come through to the other side. Bleed-through on the other hand, happens when the ink seeps in to the other side and interferes with the text on the front page. Both issues make character recognition very challenging, specially bleed-through.

Figure 6a shows a low contrast image and its binarized one. The current binarization methods are not able to recover the text properly when text has low contrast. Figure 6b presents another example with bleed though present in the image. As one can see, the method was not able to remove the bleed through completely. Figure 6c shows an example of an RGB image and its binarized one. As you can see, the method is not performing well over texts with orange color. Thus to address these issues we need to develop a method that would take them into account.

6.5 OCR performance evaluation

One of the main purpose of document image enhancement is to enhance character recognition methods or OCR to facilitate automated document analysis. Currently there is no document image test dataset with the extracted ground truth text so that could be utilized to evaluate document images enhancement methods in terms of OCR improvement. Current methods either ignore to evaluate their methods in terms of OCR, or show the OCR improvement only on a few images which is not sufficient to prove the practicality of their methods in the wild. This calls for a separate study to collect such dataset and benchmark current methods against this test dataset.

7 Conclusion

In this paper we reviewed deep learning based methods for six different document image enhancement tasks, including binarization, debluring, denoising, defading, watermark removal, and shadow removal. We also summarized datasets used for these tasks along with the metrics used to evaluate the performance of these methods. We discussed the features, challenges, advantages and disadvantages of the deep learning based document image enhancement methods.

We also discussed open problems in this area and identified multiple important tasks that have received little to no attention. These tasks are over-exposure/under-exposure correction, defading, and super-resolution. Over-exposure problem usually occurs when the imaging device captures too much light or glare due to reflection, and under-exposure occurs when the lighting condition is poor and the captured image becomes dark and hard to read. Fading could happen due to sunlight, aging, and being washed out, *etc*. Low-resolution document images need to be super-resolved to enhance their visual quality and more importantly make small text more legible. Enhancing the document image resolution is more challenging when noise and artifacts are present in the document image. Such images are often hard to read and the low legibility affects the performance of character recognition techniques. The above-explained tasks have received little attention and they are far from solved.

Binarization task has received a great deal of attention over the past years, however, these methods underperform in multiple scenarios. For example, when the image has low contrast or multiple artifacts *e.g.,* stamp, signature, ghosting or bleed-through are present. Ghosting and bleed-through occur when the text from the other side of the document can be seen or ink seeps in to the other side of the document. These artifacts are challenging to remove and effective methods are needed to address and resolve these problems properly.

Current document image enhancement methods mainly focus on improving the visual quality of the images. While this is an important aspect, the performance of these methods for automatic document analysis problems, *e.g.,* character recognition, is largely ignored. Thus there is an emerging need to develop methods that can jointly enhance the visual quality and OCR performance. The OCR performance needs to be evaluated over a larger test dataset, and not just over a few samples as was done in the literature.

All that said, current methods target only one problem, *e.g.,* debluring, at a time, but in reality a document image can have multiple issues at the same time. For example, a document image could be blurry, faded and noisy. To the best of our knowledge, currently these is no method that can tackle multiple issues in a single image at the same time.

References

[1] AFIFI, M., DERPANIS, K. G., OMMER, B., AND BROWN, M. S. Learning to correct overexposed and underexposed photos. *arXiv preprint arXiv:2003.11596 13* (2020).

[2] ANVARI, Z., AND ATHITSOS, V. A pipeline for automated face dataset creation from unlabeled images. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments* (2019), pp. 227–235.

[3] ANVARI, Z., AND ATHITSOS, V. Enhanced cyclegan dehazing network. In *VISIGRAPP (4: VISAPP)* (2021), pp. 193–202.

[4] ARLAZAROV, V. V., BULATOV, K. B., CHERNOV, T. S., AND ARLAZAROV, V. L. Midv-500: a dataset for identity document analysis and recognition on mobile devices in video stream.

[5] CAI, J., GU, S., AND ZHANG, L. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing 27*, 4 (2018), 2049–2062.

[6] CALVO-ZARAGOZA, J., AND GALLEGO, A.-J. A selectional auto-encoder approach for document image binarization. *Pattern Recognition 86* (2019), 37–47.

[7] CHEN, X., HE, X., YANG, J., AND WU, Q. An effective document image deblurring algorithm. In *CVPR 2011* (2011), IEEE, pp. 369–376.

[8] CHU, M., XIE, Y., LEAL-TAIXÉ, L., AND THUEREY, N. Temporally coherent gans for video super-resolution (tecogan). *arXiv preprint arXiv:1811.09393 1*, 2 (2018), 3.

[9] DENG, F., WU, Z., LU, Z., AND BROWN, M. S. Binarizationshop: a user-assisted software suite for converting old documents to black-and-white. In *Proceedings of the 10th annual joint conference on Digital libraries* (2010), pp. 255–258.

[10] DEY, S., AND JAWANPURIA, P. Light-weight document image cleanup using perceptual loss. *arXiv preprint arXiv:2105.09076* (2021).

[11] DONG, C., LOY, C. C., HE, K., AND TANG, X. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence 38*, 2 (2015), 295–307.

[12] DUA, D., AND GRAFF, C. UCI machine learning repository, 2017.

[13] GANGEH, M. J., PLATA, M., MOTAHARI, H., AND DUFFY, N. P. End-to-end unsupervised document image blind denoising. *arXiv preprint arXiv:2105.09437* (2021).

[14] GANGEH, M. J., TIYYAGURA, S. R., DASARATHA, S. V., MOTAHARI, H., AND DUFFY, N. P. Document enhancement system using auto-encoders. In *Workshop on Document Intelligence at NeurIPS 2019* (2019).

[15] GATOS, B., NTIROGIANNIS, K., AND PRATIKAKIS, I. Icdar 2009 document image binarization contest (dibco 2009). In *2009 10th International conference on document analysis and recognition* (2009), IEEE, pp. 1375–1382.

[16] GU, S., ZUO, W., GUO, S., CHEN, Y., CHEN, C., AND ZHANG, L. Learning dynamic guidance for depth image enhancement. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 3769–3778.

[17] GUO, C., LI, C., GUO, J., LOY, C. C., HOU, J., KWONG, S., AND CONG, R. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 1780–1789.

[18] GUO, X., LI, Y., AND LING, H. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on image processing 26*, 2 (2016), 982–993.

[19] HE, K., ZHANG, X., REN, S., AND SUN, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 770–778.

[20] HE, S., AND SCHOMAKER, L. Deepotsu: Document enhancement and binarization using iterative deep learning. *Pattern recognition 91* (2019), 379–390.

[21] HEDJAM, R., NAFCHI, H. Z., MOGHADDAM, R. F., KALACSKA, M., AND CHERIET, M. Icdar 2015 contest on multispectral text extraction (ms-tex 2015). In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (2015), IEEE, pp. 1181–1185.

[22] HOWE, N. R. Document binarization with automatic parameter tuning. *International journal on document analysis and recognition (ijdar) 16*, 3 (2013), 247–258.

[23] HRADIŠ, M., KOTERA, J., ZEMCIK, P., AND ŠROUBEK, F. Convolutional neural networks for direct text deblurring. In *Proceedings of BMVC* (2015), vol. 10.

[24] IOFFE, S., AND SZEGEDY, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning* (2015), PMLR, pp. 448–456.

[25] JEMNI, S. K., SOUIBGUI, M. A., KESSENTINI, Y., AND FORNÉS, A. Enhance to read better: An improved generative adversarial network for handwritten document image enhancement. *arXiv preprint arXiv:2105.12710* (2021).

[26] JIANG, Y., GONG, X., LIU, D., CHENG, Y., FANG, C., SHEN, X., YANG, J., ZHOU, P., AND WANG, Z. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing 30* (2021), 2340–2349.

[27] JO, Y., OH, S. W., KANG, J., AND KIM, S. J. Deep video super-resolution network using dynamic upsampling filters without explicit motion compensation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 3224–3232.

[28] KIM, J., LEE, J. K., AND LEE, K. M. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 1637–1645.

[29] KLIGLER, N., KATZ, S., AND TAL, A. Document enhancement using visibility detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 2374–2382.

[30] LAI, W.-S., HUANG, J.-B., AHUJA, N., AND YANG, M.-H. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 624–632.

[31] LEDIG, C., THEIS, L., HUSZÁR, F., CABALLERO, J., CUNNINGHAM, A., ACOSTA, A., AITKEN, A., TEJANI, A., TOTZ, J., WANG, Z., ET AL. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 4681–4690.

[32] LEWIS, D., AGAM, G., ARGAMON, S., FRIEDER, O., GROSSMAN, D., AND HEARD, J. Building a test collection for complex document information processing. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval* (2006), pp. 665–666.

[33] LI, D., WU, Y., AND ZHOU, Y. Sauvolanet: Learning adaptive sauvola network for degraded document binarization. *arXiv preprint arXiv:2105.05521* (2021).

[34] LI, X., ZHANG, B., LIAO, J., AND SANDER, P. V. Document rectification and illumination correction using a patch-based cnn. *ACM Transactions on Graphics (TOG) 38*, 6 (2019), 1–11.

[35] LIN, W.-A., CHEN, J.-C., CASTILLO, C. D., AND CHELLAPPA, R. Deep density clustering of unconstrained faces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 8128–8137.

[36] LIN, Y.-H., CHEN, W.-C., AND CHUANG, Y.-Y. Bedsr-net: A deep shadow removal network from a single document image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 12905–12914.

[37] LIU, W., ANGUELOV, D., ERHAN, D., SZEGEDY, C., REED, S., FU, C.-Y., AND BERG, A. C. Ssd: Single shot multibox detector. In *European conference on computer vision* (2016), Springer, pp. 21–37.

[38] LONG, J., SHELHAMER, E., AND DARRELL, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 3431–3440.

[39] LU, H., KOT, A. C., AND SHI, Y. Q. Distance-reciprocal distortion measure for binary document images. *IEEE Signal Processing Letters 11*, 2 (2004), 228–231.

[40] MAO, X., SHEN, C., AND YANG, Y.-B. Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections. *Advances in neural information processing systems 29* (2016), 2802–2810.

[41] MESQUITA, R. G., MELLO, C. A., AND ALMEIDA, L. A new thresholding algorithm for document images based on the perception of objects by distance. *Integrated Computer-Aided Engineering 21*, 2 (2014), 133–146.

[42] MICHALAK, H., AND OKARMA, K. Robust combined binarization method of non-uniformly illuminated document images for alphanumerical character recognition. *Sensors 20*, 10 (2020), 2914.

[43] NAFCHI, H. Z., AYATOLLAHI, S. M., MOGHADDAM, R. F., AND CHERIET, M. An efficient ground truthing tool for binarization of historical manuscripts. In *2013 12th International Conference on Document Analysis and Recognition* (2013), IEEE, pp. 807–811.

[44] NAYEF, N., LUQMAN, M. M., PRUM, S., ESKENAZI, S., CHAZALON, J., AND OGIER, J.-M. Smartdoc-qa: A dataset for quality assessment of smartphone captured document images-single and multiple distortions. In *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (2015), IEEE, pp. 1231–1235.

[45] NTIROGIANNIS, K., GATOS, B., AND PRATIKAKIS, I. Performance evaluation methodology for historical document image binarization. *IEEE Transactions on Image Processing 22*, 2 (2012), 595–609.

[46] NTIROGIANNIS, K., GATOS, B., AND PRATIKAKIS, I. Icfhr2014 competition on handwritten document image binarization (h-dibco 2014). In *2014 14th International conference on frontiers in handwriting recognition* (2014), IEEE, pp. 809–813.

[47] PANDEY, R. K., AND RAMAKRISHNAN, A. Language independent single document image super-resolution using cnn for improved recognition. *arXiv preprint arXiv:1701.08835* (2017).

[48] PENG, X., AND WANG, C. Building super-resolution image generator for ocr accuracy improvement. In *International Workshop on Document Analysis Systems* (2020), Springer, pp. 145–160.

[49] PRATIKAKIS, I., GATOS, B., AND NTIROGIANNIS, K. H-dibco 2010-handwritten document image binarization competition. In *2010 12th International Conference on Frontiers in Handwriting Recognition* (2010), IEEE, pp. 727–732.

[50] PRATIKAKIS, I., GATOS, B., AND NTIROGIANNIS, K. Icdar 2011 document image binarization contest (dibco 2011). In *2011 International Conference on Document Analysis and Recognition* (2011), pp. 1506–1510.

[51] PRATIKAKIS, I., GATOS, B., AND NTIROGIANNIS, K. Icfhr 2012 competition on handwritten document image binarization (h-dibco 2012). In *2012 international conference on frontiers in handwriting recognition* (2012), IEEE, pp. 817–822.

[52] PRATIKAKIS, I., GATOS, B., AND NTIROGIANNIS, K. Icdar 2013 document image binarization contest (dibco 2013). In *2013 12th International Conference on Document Analysis and Recognition* (2013), IEEE, pp. 1471–1476.

[53] PRATIKAKIS, I., ZAGORI, K., KADDAS, P., AND GATOS, B. Icfhr 2018 competition on handwritten document image binarization (h-dibco 2018). In *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (2018), pp. 489–493.

[54] PRATIKAKIS, I., ZAGORIS, K., BARLAS, G., AND GATOS, B. Icfhr2016 handwritten document image binarization contest (h-dibco 2016). In *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)* (2016), IEEE, pp. 619–623.

[55] PRATIKAKIS, I., ZAGORIS, K., BARLAS, G., AND GATOS, B. Icdar2017 competition on document image binarization (dibco 2017). In *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)* (2017), vol. 1, IEEE, pp. 1395–1403.

[56] REDMON, J., DIVVALA, S., GIRSHICK, R., AND FARHADI, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016), pp. 779–788.

[57] REN, W., LIU, S., MA, L., XU, Q., XU, X., CAO, X., DU, J., AND YANG, M.-H. Low-light image enhancement via a deep hybrid network. *IEEE Transactions on Image Processing 28*, 9 (2019), 4364–4375.

[58] SAUVOLA, J., AND PIETIKÄINEN, M. Adaptive document image binarization. *Pattern recognition 33*, 2 (2000), 225–236.

[59] SCHROFF, F., KALENICHENKO, D., AND PHILBIN, J. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2015), pp. 815–823.

[60] SHARMA, M., VERMA, A., AND VIG, L. Learning to clean: A gan perspective. In *Asian Conference on Computer Vision* (2018), Springer, pp. 174–185.

[61] SOUIBGUI, M. A., AND KESSENTINI, Y. De-gan: A conditional generative adversarial network for document enhancement. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).

[62] 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.

[63] SU, B., LU, S., AND TAN, C. L. Robust document image binarization technique for degraded document images. *IEEE transactions on image processing 22*, 4 (2012), 1408–1417.

[64] TAI, Y., YANG, J., AND LIU, X. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2017), pp. 3147–3155.

[65] TAO, X., GAO, H., LIAO, R., WANG, J., AND JIA, J. Detail-revealing deep video super-resolution. In *Proceedings of the IEEE International Conference on Computer Vision* (2017), pp. 4472–4480.

[66] TENSMEYER, C., AND MARTINEZ, T. Document image binarization with fully convolutional neural networks. In *2017 14th IAPR international conference on document analysis and recognition (ICDAR)* (2017), vol. 1, IEEE, pp. 99–104.

[67] TRANSLATION, S. M. Sixth workshop on statistical machine translation.

[68] VAN DER ZANT, T., SCHOMAKER, L., AND HAAK, K. Handwritten-word spotting using biologically inspired features. *Ieee transactions on pattern analysis and machine intelligence 30*, 11 (2008), 1945–1957.

[69] WANG, R., ZHANG, Q., FU, C.-W., SHEN, X., ZHENG, W.-S., AND JIA, J. Underexposed photo enhancement using deep illumination estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 6849–6857.

[70] 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.

[71] WANG, X., YU, K., WU, S., GU, J., LIU, Y., DONG, C., QIAO, Y., AND CHANGE LOY, C. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops* (2018), pp. 0–0.

[72] WANG, Z., BOVIK, A. C., SHEIKH, H. R., AND SIMONCELLI, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing 13*, 4 (2004), 600–612.

[73] XU, X., SUN, D., PAN, J., ZHANG, Y., PFISTER, H., AND YANG, M.-H. Learning to super-resolve blurry face and text images. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 251–260.

[74] ZHAO, G., LIU, J., JIANG, J., GUAN, H., AND WEN, J.-R. Skip-connected deep convolutional autoencoder for restoration of document images. In *2018 24th International Conference on Pattern Recognition (ICPR)* (2018), IEEE, pp. 2935–2940.