Table of Contents

1 Part 1

1) Implement the Ostu global thresholding algorithm for binarizing the sample text images and feed the binarized images to the OCR software to evaluate the OCR accuracy. Discuss any problems with the Otsu global thresholding algorithm.

We have worked on both of the sample text images downloaded from the project website on NTULearn.

1.1 Sample01

We have plotted the histogram of the image. The x-axis shows the pixel's brightness, and the y-axis shows the number of pixels that belong to a particular brightness. We can see that the image has a very bad contrast from the histogram as many pixels belong to the different brightness. Moreover, there are two peaks located at very different brightness. One is located at around 70, which is quite dark, whereas the other one is located at around 210, which is quite bright. This will prove to be a problem in Otsu global thresholding as the poor contrast would make it difficult to enhance the image, which in the end result in the OCR's difficulty in recognizing the text in the image. Throughout this project, we will be implementing algorithms which aim to improve the distribution of the histogram.

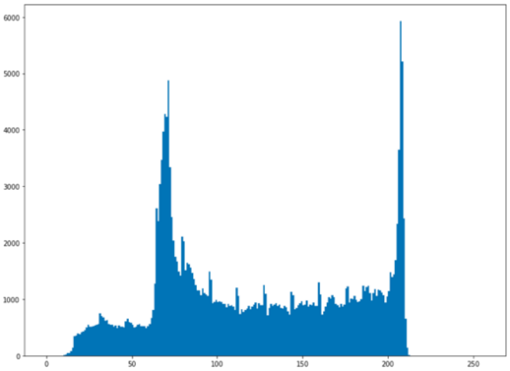
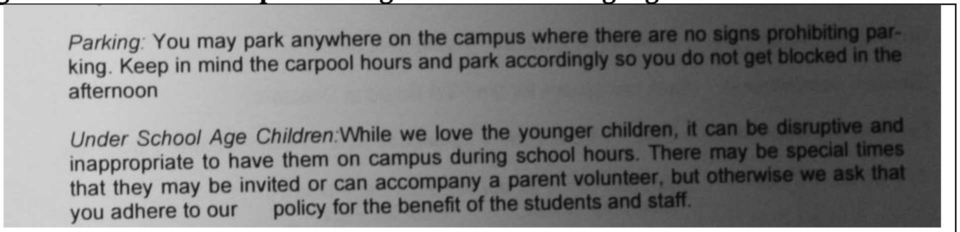
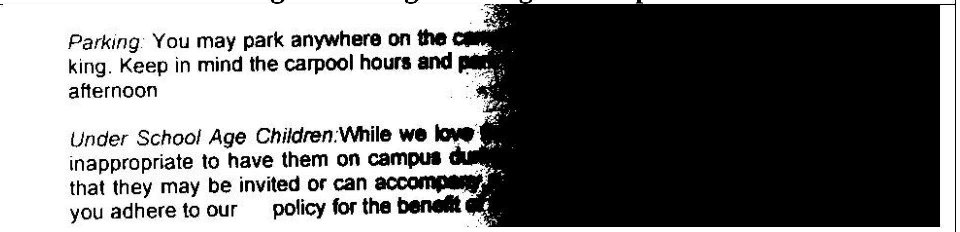


Figure 1: Histogram of Sample01

Table 1: Sample01: Image before and after implementing Otsu thresholding algorithm

Original Image	 *Parking:* You may park anywhere on the campus where there are no signs prohibiting par-king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon *Under School Age Children:* While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 2: Original Image of Sample01**
Image after implementing Ostu thresholding algorithm, threshold=127	 *Parking:* You may park anywhere on the king. Keep in mind the carpool hours and afternoon *Under School Age Children:* While we love inappropriate to have them on campus that they may be invited or can accompany you adhere to our policy for the benefit **Figure 3: Sample01 Image after implementing Otsu thresholding algorithm**

Text predicted by OCR software:

```
Parking You may park anywhere on the &
king. Keep in mind the carpool hours and g
afternoon

Under School Age Children:While we love'
inappropriate to have them on campus

that they may be invited or can accompesy. i
you adhere to our  policy for the benefit of}
```

Number of words correctly identified: 45 words (49.5%)

Results and observations:

The performance of the otsu thresholding algorithm is not very good. It classifies a pixel as either black or white according to whether it is greater than or lower than the otsu threshold value of 127. We can see that the otsu thresholding algorithm divides the image into two distinct parts from the results. The left part shows a clear text, whereas the right part is black. This is because the brightness of the right-hand side is too low. Hence the algorithm classifies the area as black. This shows the limitation of the otsu global thresholding algorithm as it applies the threshold to the whole image even though the image has varying levels of brightness. Due to the poor quality of the output of the otsu thresholding algorithm, the tesseractOCR cannot read the right-hand side of the text. It can still read the text's left-hand side as it is still bright enough, and the text is clear enough.

1.2 Sample02

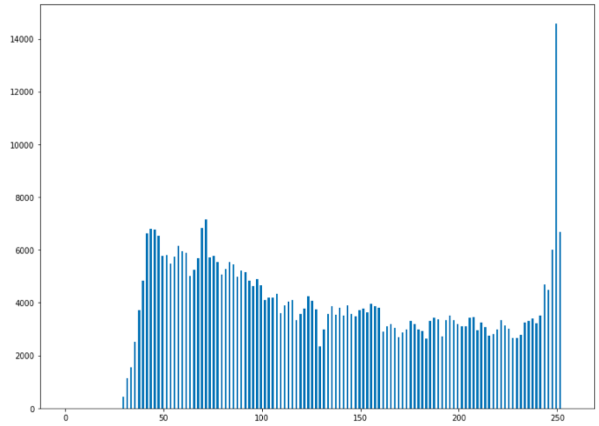
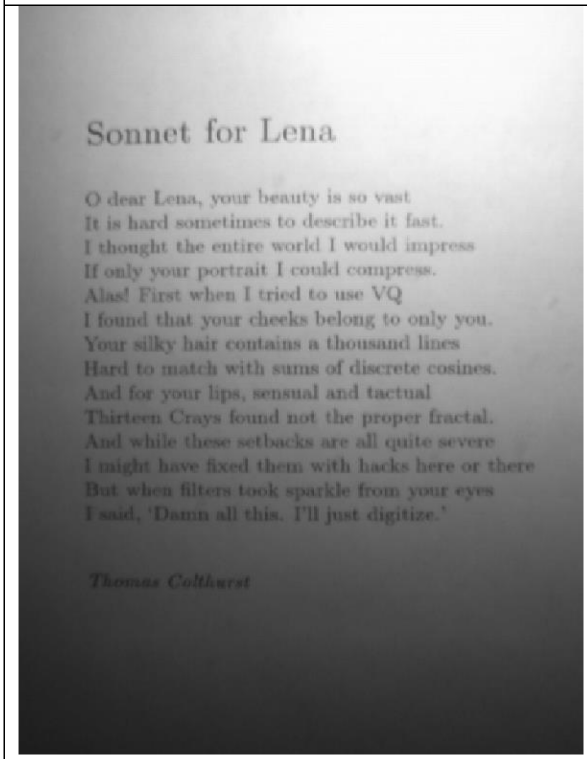
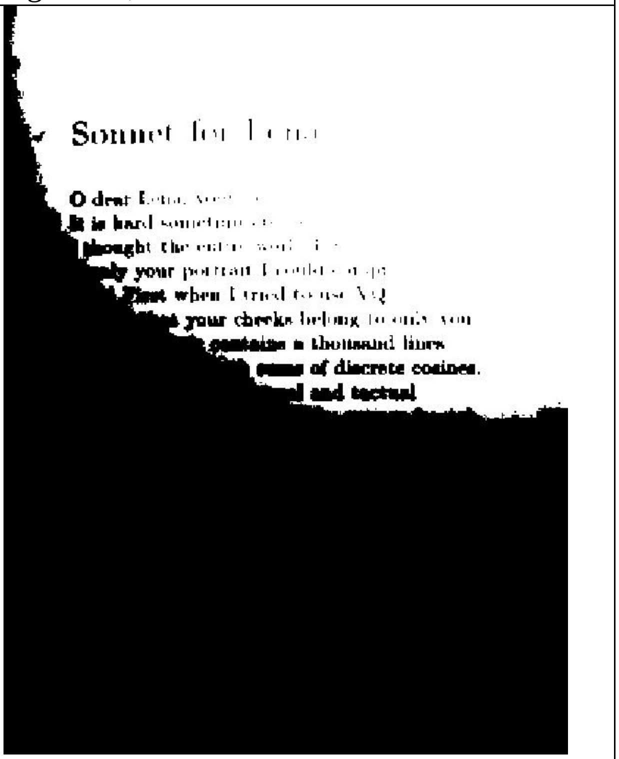


Figure 4: Histogram of Sample02

This is the histogram of the second image. Similar to Sample01, the image has a terrible contrast as many pixels belong to different brightness, and there are also two peaks located at very different brightness, in this case. Compared to the Sample01, the image has a worse brightness contrast because the range of the pixels' brightness is greater than Sample01. There is also very similar number of pixels belonging to each brightness level apart from the significant peak at the most right side of the histogram. Therefore, we predict that the image after applying Otsu global thresholding will be worse for Sample02.

Table 2: Sample02: Image before and after implementing Otsu thresholding algorithm

Original Image	Image after implementing Ostu thresholding algorithm, threshold=142
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' *Thomas Colthurst* **Figure 5: Original Image of sample02**	 **Figure 6: Sample02 Image after implementing Otsu thresholding algorithm**

Text predicted by OCR software:

```
Sonnet Ton 1o
```

Number of words correctly identified: 3 words (2.6%)

Results and observations:

As predicted, the image after applying Otsu global thresholding will be worse for Sample02. This is due to the poorer contrast of the image as mentioned above. Since the top right of the image is much brighter than the pixels of Sample01, the threshold value of Sample02 is higher than Sample01. Since the outputted image after applying Otsu global thresholding is bad, the OCR software can only predict the first word of the text "Sonnet", which is the clearest word in the whole image, correctly.

2 Part 2

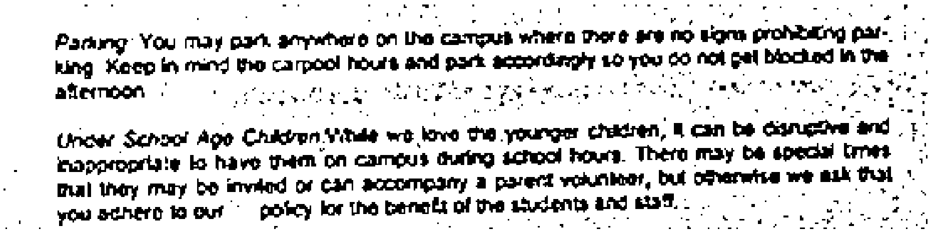
2) Design your own algorithms to address the problem of Otsu global thresholding algorithm, and evaluate OCR accuracy for the binary images as produced by your algorithms. You may explore different approaches such as adaptive thresholding, image enhancement, etc., and the target is to achieve the best OCR accuracy.

We have tried many different approaches to improve OCR accuracy. These are a list of the approaches that we have used in the order of increasing accuracy:
1. Adaptive Mean Thresholding: threshold value is the mean of the neighborhood area.
2. Adaptive Gaussian Thresholding: threshold value is the weighted sum of neighborhood values where weights are a gaussian window.
3. Increasing brightness
4. CHALE Histogram equalization
5. CLOVA

2.1 Adaptive Mean Thresholding

2.1.1 Sample01

Table 3: Sample01: Image before and after implementing Adaptive Mean Thresholding algorithm

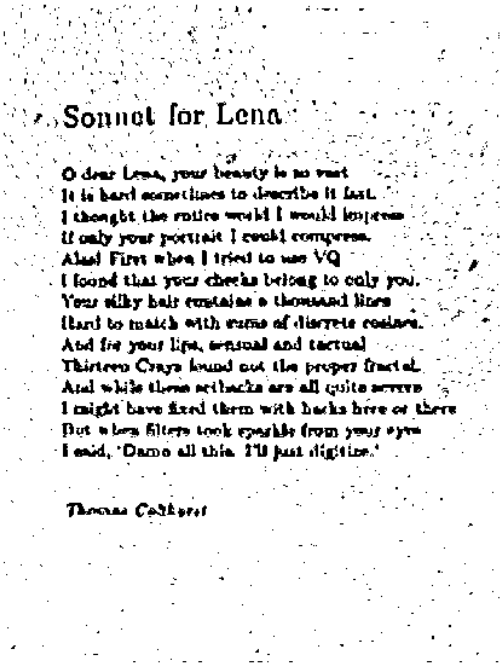
Image after implementing Adaptive Mean Thresholding	 Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon. Under School Age Children: While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 7: Sample01 Image after implementing Adaptive Mean Thresholding algorithm**

Text predicted by OCR software	`Parhing You may 5 smywerd On NS CETGAEh WARTD Do g 400 1 g ORI par |`
	`g Taco iy i) carpct Pt 3 park Sz Sehy 3070 60 (A ot blockad I D -`
	
	`atemoon 7t tggin`
	
	`you sener`

Number of words correctly identified: 2 words (2.2%)

2.1.2 Sample02

Table 4: Sample02: Image before and after implementing Adaptive Mean Thresholding algorithm

Image after implementing Adaptive Mean Thresholding	Text predicted by OCR software
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found out the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 8: Sample02 Image after implementing Adaptive Mean Thresholding algorithm**	z,.Sounsfi [Dl"'L'QIll\." : v odmlau.;vut-cwtyl-mv-(* 10 44 bard scemeclines to drecribe Dat. - 1 thoaght Lt rotice weabd | woudd bepree 12 oaly your porvieit | coah] comperss. L . Adusd Fires whea | 15t 1o wam VQ - e < € foand that yous cheehs beivag to caly you. Vour slky bals eoutalan s thowand lizes fhand to toaich with varms of diseote codare. Abd fi yous lipw, ernisoal and tacsadd o Thineen Crazs knuad ok the proger fiactel Asad while tlerme acitcka ars all uite sovove - 1 night bave fxed them with hecks heve of there Dt » e Blters tonk eyaskby (rom yomg oy Fonid, Dame ol thia T just igion. Thocas Codbprst

Number of words correctly identified: 13 words (11.2%)

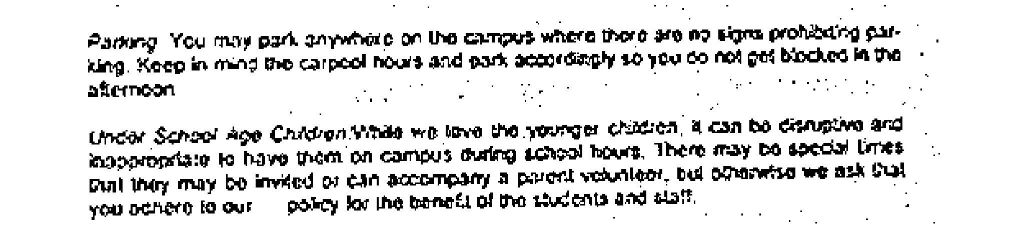
Results and observations:

From the two images above, we can see that using adaptive mean thresholding worsens the image. The text predicted by the OCR software is unintelligible or meaningless. This is because adaptive mean thresholding adds a lot more noise to the images, resulting in the OCR's inability to recognize the text in these images. Even to the naked eye, we are unable to read the text in these images. This shows the severe limitation of adaptive mean thresholding in improving the quality of the text.

2.2 Adaptive Gaussian Thresholding

2.2.1 Sample01

Table 5: Sample01: Image before and after implementing Adaptive Gaussian Thresholding algorithm

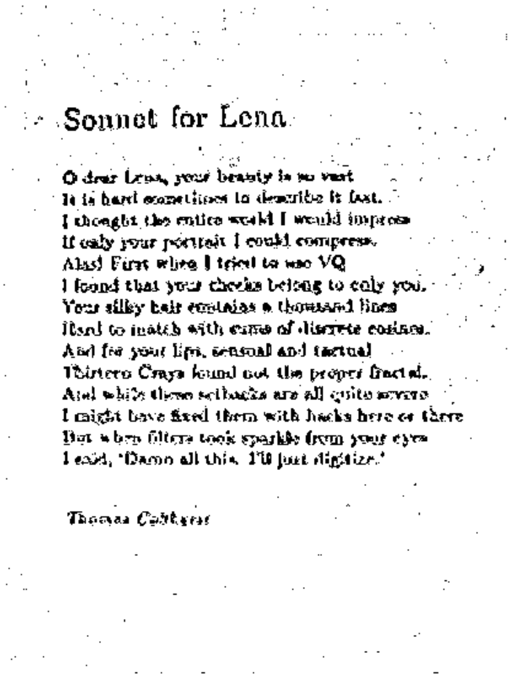
Image after implementing Adaptive Gaussian Thresholding	 *Parking* You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon *Under School Age Children* While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. Figure 9: Sample01 Image after implementing Adaptive Gaussian Thresholding algorithm

Text predicted by OCR software	Paing You may pan I On N6 CHTRS Where (horg 30 80 Figrs oy gt King. Koep tn ming o carpeoi hoors and D ARG 30 you 60 A01 g0 bockes I ntha - atiemoon 5 , o . Uncar Senool A CRADON VIS wD Iove the yunger CRICIEN, 4 C3N 80 Cinive 33 Tnapproprary 1o have e BN CaMB' Ouig 4CH53] bow, There M3y B0 speod Lm es Dl thay may bo ik 0f €30 SCOMEINY 3 e vekAledr. bul 0aswtse w 82K D13t o BERED [0 ur POFCY K IhG BINSE Bf the leAls 2 4Lt

Number of words correctly identified: 6 words (6.6%)

2.2.2 Sample02

Table 6: Sample02: Image before and after implementing Adaptive Gaussian Thresholding algorithm

Image after implementing Adaptive Gaussian Thresholding	Text predicted by OCR software
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 10: Sample02 Image after implementing Adaptive Gaussian Thresholding algorithm**	+.Somnet for Lena O drar Urox, yos beanty b wo vest - 10 14 bared scamettions Lo demritz bt Daat. 1 thcaght th entico scaM [weuld impros 18 caly your porredt | coub) eompeess. Alast Firss wlira | teiedt ta 00 VQ 1 foand that your cheris bezag o caly you. Vo silley bals eemiaing o tomand Hoes itand 6o inated #ith eams of disrece cosian. Aisd fea yout lips, seasoad and tarsual Thirtees Crays bound ot Um prapes ractad. Al whils dlimo selbacks are 11 grits mvere 1 might bave Sxed them with hasks her ce thers B wbrn Flters took arkds e yous eyen Lead, Bamn all thin, 11 pust digitize. Thanea Catbane

Number of words correctly identified: 16 words (13.8%)

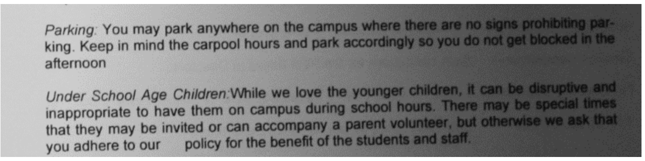
Results and observations:

Similar to adaptive mean thresholding, adaptive gaussian thresholding is also limited in improving the clarity of the image. This is also because adaptive gaussian thresholding produces much noise, hindering the OCR's ability to recognize the text in the images. The text predicted by the OCR software for the two images is unintelligible or meaningless.

2.3 Increasing brightness

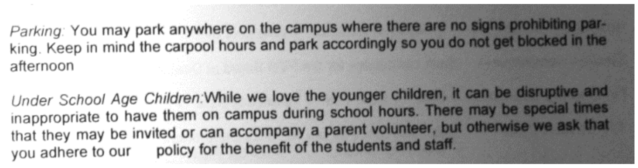
2.3.1 Sample01

Table 7: Sample01: Original + otsu algorithm

Original + otsu algorithm	
Original Image	 *Parking:* You may park anywhere on the campus where there are no signs prohibiting par-king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon *Under School Age Children:*While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 11: Original Sample01 Image**

Text predicted by OCR software	`Parking' You may park anywhere on the &`
	`king. Keep in mind the carpool hours and pés`
	`afternoon`
	
	`Under School Age Children:While we love'`
	`inappropriate to have them on campus @`
	`that they may be invited of can acCompanY, i`
	`you adhere to our  policy for the benefit of}`

Table 8: Sample01: Increasing brightness(1.5 times brighter) + otsu algorithm

Increasing brightness(1.5 times brighter) + otsu algorithm	
Image after increasing brightness (1.5x)	 *Parking:* You may park anywhere on the campus where there are no signs prohibiting par-king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon *Under School Age Children:*While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 12: Sample01 Image after increasing brightness (1.5x)**

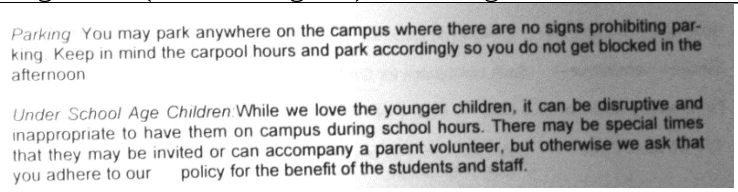
Text predicted by OCR software	`Parking' You may park anywhere on the campus`
	`king. Keep in mind the carpool hours and park`
	`afternoon`
	
	`Under School Age Children:While we love the.`
	`inappropriate to have them on campus during`
	`that they may be invited or can accompany a P`
	`you adhere to our  policy for the benefit of the`

Note: the words highlighted in green are the additional words that the OCR can identify compared to the original image

Results and observations:

We have highlighted the additional words that the OCR software can identify in green when the brightness increased 1.5 times. We can see an improvement in the coverage of words of the OCR software.

Table 9: Sample01: Increasing brightness(2 times brighter) + otsu algorithm

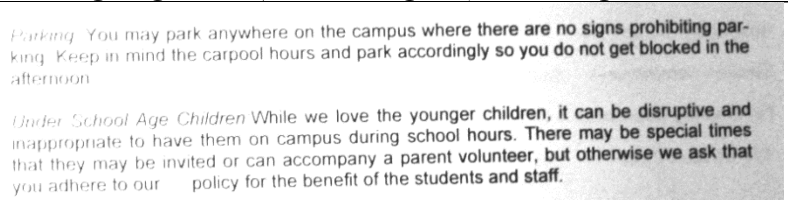
Increasing brightness (2 times brighter) + otsu algorithm	
Image after increasing brightness (2x)	 *Parking:* You may park anywhere on the campus where there are no signs prohibiting par-king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon *Under School Age Children:*While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 13: Sample01 Image after increasing brightness (2x)**

Text predicted by OCR software	Parking You may park anywhere on the campus whare king Keep in mind the carpool hours and park accordingly afternoon 3 Under School Age Children:While we love the younger @ inappropriate to have them on campus during school that they may be invited or can accompany a parent N you adhere to our policy for the benefit of the students

Note: the words highlighted in green are the additional words correctly identified compared to the image after increasing brightness(1.5x). In contrast, the words highlighted in red are the additional words incorrectly identified compared to the image after increasing brightness (2x).

Results and observations: As expected, the coverage of words identified by the OCR software has increased when the image's brightness increases. However, we can also observe that increasing brightness is starting to negatively affect the image through the addition of noise in the image. This is evident through the presence of the additional words that are incorrectly spelled.

Table 10: Sample01: Increasing brightness(3 times brighter) + otsu algorithm

Increasing brightness (3 times brighter) + otsu algorithm	
Image after increasing brightness (3x)	 Parking You may park anywhere on the campus where there are no signs prohibiting par- king Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 14: Sample01 image after increasing brightness (3x)**
Text predicted by OCR software	ioamewag {ou sy park anywhere on the campus where there are no signe s Wiy Weep i mind the carpoot hours and park accordingly so you do not get blocked In the L aAtternonn sindes dvaol Aye Children While we love the younger children, it can be disnptve.arel . Japproptiate 1o have them on campus during school hours. There may b e speciel tmes ~ Wiat they may be invited or can accompany a parent volunteer, but ot herwise we aek-dk: ¥ you adhere to our policy for the benefit of the students and staff.

Note: the words highlighted in green are the additional words correctly identified compared to the image after increasing brightness(2x). In contrast, the words highlighted in red are the words incorrectly identified compared to the image after increasing brightness (3x).

Results and observations: The coverage of words identified by the OCR software has increased. However, the increase in noise is a lot more significant as more words are incorrectly identified both at the left of the image and also the right of the image.

Optimal brightness for left and right respectively based on above findings

Overall, we have concluded that the OCR software can only read the left half of the image correctly if the brightness is less than 2x. However, the right half of the image can only be read by the OCR software if the image's brightness is more than 2x.

Splitting the image through median filtering

Therefore, we decided to cut the image into two parts and change the brightness of each part accordingly, such that the brightness of the first half is 2x whereas the brightness of the second half is 3x and then merge these two separate parts back together to form the original image.

To determine the boundary to split the image, global filtering is used first to find the image's bright and dark regions. A very large median filter is used to make the bright and dark regions more distinct. The image below shows the median filtering result. Afterward, the points of the line of the boundary between the white and the black are calculated, and the mean of the points will determine the midpoint to split the image.

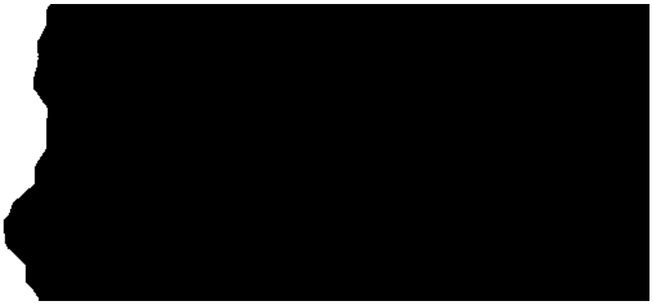
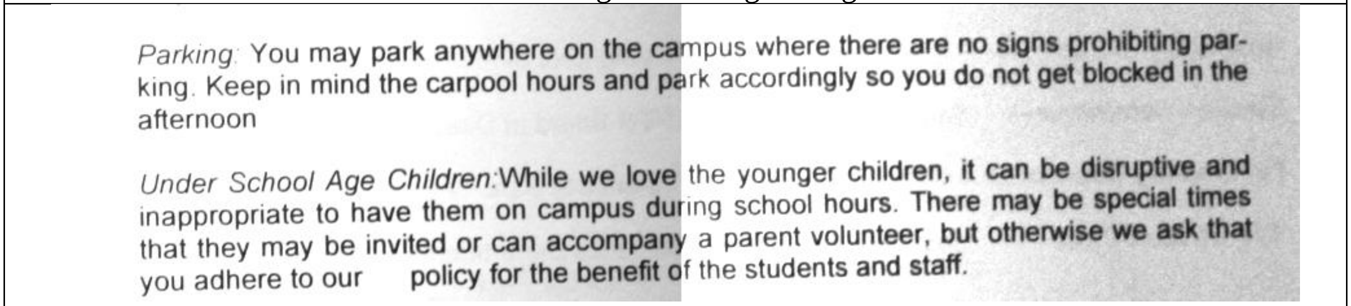


Figure 15: Median filtering result of Sample01

Table 11: Sample01: Combining left and right images

Combining left and right images
 *Parking:* You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon *Under School Age Children:* While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 16: Sample01 image after combining the left and right splitted images**

`Parking’ You may park anywhere on the campus where there are no signs prohibiting par-`
`king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the`
`afternoon`
 	
`Under School Age Children:While we love the younger children, it can be disruptive and`
`inappropriate to have them on campus during school hours. There may be special times`
`that they may be invited or can accompany a parent volunteer, but otherwise we ask that`
`you adhere to our  policy for the benefit of the students and staff.`

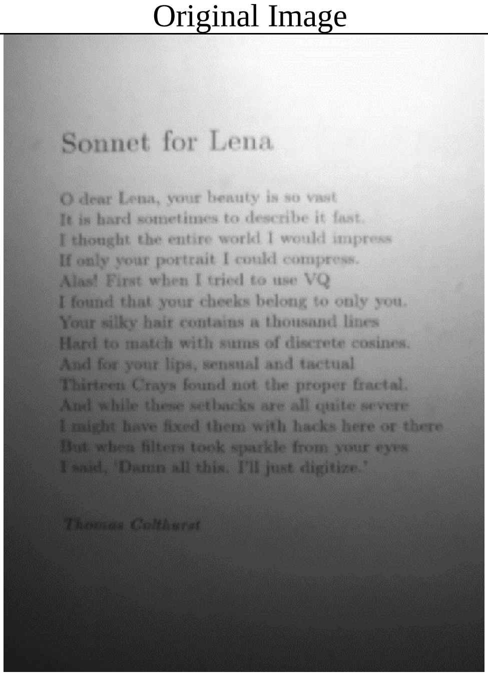
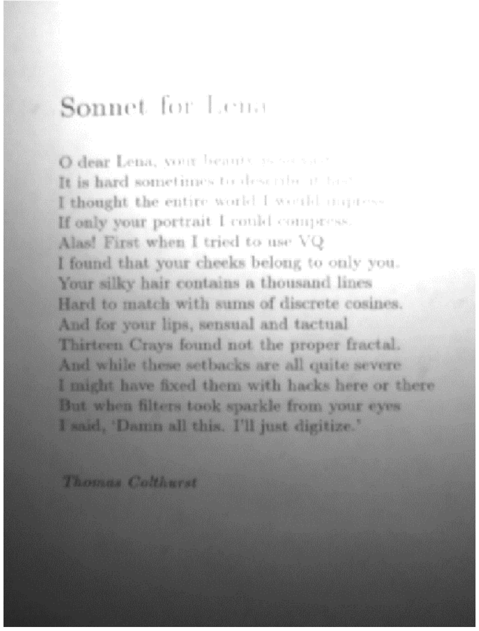
Number of words correctly identified: 91 words (100%)

Results and observations:

We can see that the all the words are correctly identified. The only error is 1 punctuation error where ":" is incorrectly identified as " ' " . Overall, increasing brightness did a great job at improving the clarity of the image for the Tesseract OCR to recognize the text.

2.3.2 Sample02

Table 12: Sample02: Increasing brightness (1.5 times brighter) + otsu algorithm

Increasing brightness (1.5 times brighter) + otsu algorithm	
Original Image	Image after increasing brightness (1.5x)
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 17: Original Sample02 image**	 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 18: Sample02 image after increasing brightness (1.5x)**
Text predicted by OCR software	
Sonnet lon 1o Odenr Bt v i Ward s bt Che : FOUE P et Daenila s when [ried e 3y yaur cheeke Delomg 1o oniy von w thuisaid lines and tactunl	Nennet G T~ L than Il oo e Abwa® Fiis wlo fownd tlun & o ra e il wind a1l found not the proper fractal. are all guile severe

From the two results above, we can see that even though increasing brightness positively affects the first image, the OCR result's improvement is very small. Most of the words outputted are still unintelligible or meaningless.

Splitting the image through median filtering

We have tried to split the image through median filtering, similar to what was done to sample01. The difference between Sample01 is that instead of splitting the image vertically into the left part and right part, we will be splitting horizontally into the up part and down part. Afterward, we will then change the brightness of each part accordingly.

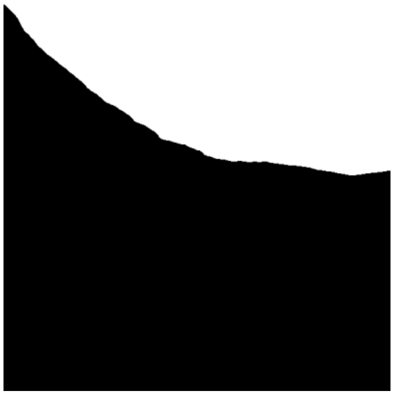
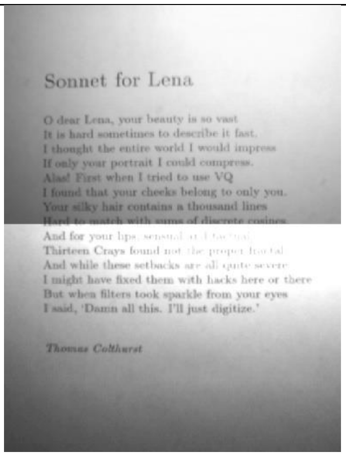


Figure 19: Mean filtering result of Sample02

Table 13: Sample02: Combining left and right images

Combining left and right images	
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' *Thomas Colthurst* **Figure 20: Sample02 image after combining the l eft and right splitted images**	et for Lena your beauty is so vasi to describe it fast entire world 1 would impress And for setbacks as anight have fixed them with hacks here or thers when filters took sparkle from your eyes all this. I'll just digitize.'

Number of words correctly identified: 37 words (32.0%)

Results and observations:

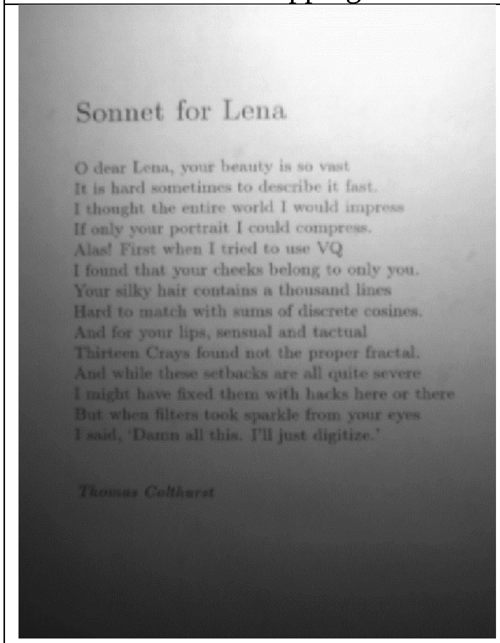
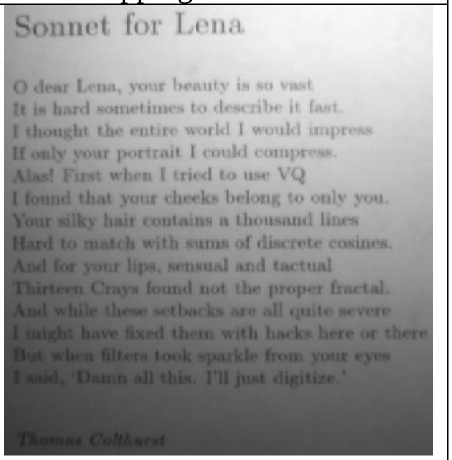
The coverage of words identified by the OCR software has increased, and only a few words are gibberish. However, the number of words identified is still very small. This may be due to the fact that the top right corner is very bright, and the bottom left corner is very dark—the extreme contrast results in the ineffectiveness of brightening the image. Therefore, we try to do some modifications to deal with this limitation and improve the image's quality, such as through cropping of the image and CHALE histogram equalization. These modifications will be mentioned below.

2.4 Further modifications for Sample02 : Cropping and CHALE Histogram equalization

Since the accuracy of the Sample01 is 100 percent, we will not be applying further modifications to the image.

We will be carrying out modifications for Sample02. Firstly, the image is cropped only to encompass the text. This is because there is an extreme contrast in the top corner and bottom corner, which made the brightness adjustments very bad even after splitting the image. Therefore, cropping reduces the area of the bright and dark corners.

Table 14: Sample02: Image before and after cropping

Before cropping	After cropping
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 21: Original Sample02 image**	 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 22: Sample02 image after cropping**

We also used CHALE Histogram equalization to sharpen the image text. CHALE stands for Contrast Limited Adaptive Histogram Equalization and operates on small regions in the image, called tiles, rather than the entire image. The neighboring tiles are then combined using bilinear interpolation to remove the artificial boundaries. This algorithm can be applied to improve the contrast of images. Therefore, CHALE Histogram equalization is very useful in our example since our image's brightness varies throughout and prevents the over-amplification of the contrast, especially for the dark regions on the lower left corners of the cropped image. (refer to Figure 6: sample02 after implementing Otsu thresholding algorithm for reference)

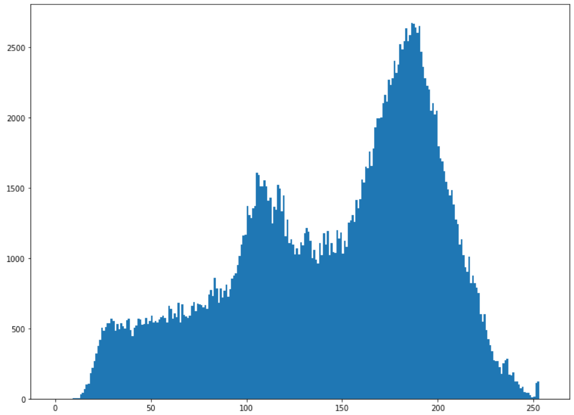
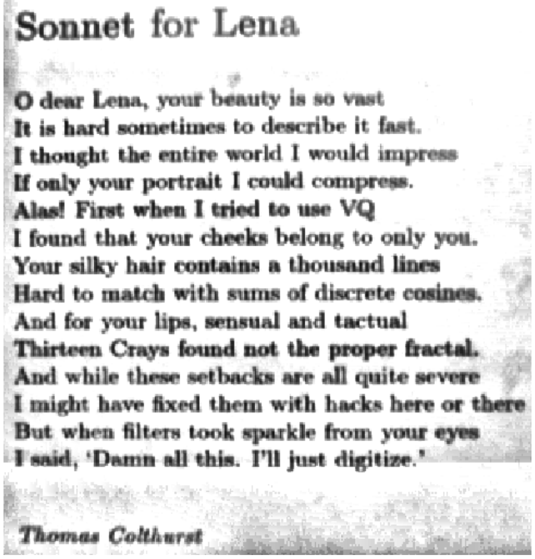


Figure 23: Histogram of Sample02 after applying CHALE histogram equalization.

Above shows the histogram of Sample02 after applying CHALE histogram equalization. The pixels' brightness is more concentrated instead of being more spread out (refer to figure 4, the histogram of original Sample02). This meant that the brightness contrast is lesser, and hence increasing the brightness of the image afterward will prove to be more effective.

After applying CHALE Histogram Equalization, we will carry out the same procedures as mentioned in Section 2.3, where we will split the image into two and increase the brightness of each sections accordingly and then combine the two separated images together.

Table 15: Sample02: Image after implementing CHALE Histogram Equalization

CHALE Histogram Equalization On Sample02	
Second image after CHALE Histogram Equalization	 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst **Figure 24: Sample02 after CHALE Histogram Equalization**
Result from OCR Testing	tSonnet for Lena O dear Lena, yout heatty is so vast Tt in hard sometimes to describe it fast, 1 thought the entire world | would impress H only your portrait | could compress. Alas! First when | tried to use VQ

	1 found that your cheeks belong to only you. Your silky hait contains s thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these sotbacks are all quite sovere 1 might have fixed them with hacks here or there But when filters took spackle from your eyes 1 wnbl, ‘Dasun all thin. T just digitise.” Thomas Colthwrst . . . .

Number of words correctly identified: 92 words (79%)

Results and observations:
The results are significantly better compared to the OCR result of figure 18. It is able to identify all sections of the text and has some spelling errors. The OCR seems to have trouble correctly identify the alphabet I, misidentifying it as “|” and even “1”. There are also some additional punctuations identified, most likely due to the image's noise.

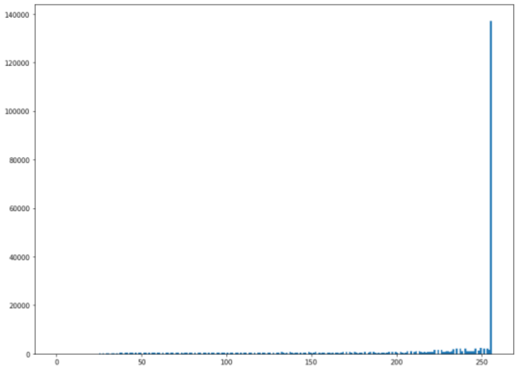


Figure 25: Histogram of Sample02 after applying CHALE histogram equalization and increasing brightness accordingly

Above shows the histogram of Sample02 after applying CHALE histogram equalization and increasing brightness accordingly. The brightness of the pixels is very concentrated at the peak which is very good.

2.5 CLOVA OCR

We have implemented the CLOVA OCR algorithm that is based on two github repositories:
1. https://github.com/clovaai/CRAFT-pytorch
2. https://github.com/clovaai/deep-text-recognition-benchmark

The two github repositories are based on two different parts, **text detection,** and **text recognition**.

For **text detection**, it is based on PyTorch implementation for CRAFT text detector. It can effectively detect text areas by exploring each character region and affinity between characters. After thresholding the character region and affinity score, it can find the minimum bounding rectangles on the binary map. This is because the affinity score groups each character into a single instance. It then bounds the texts of characters together to create the bounding box of texts. For our project, we are using a pre-trained model *craft_mlt_25k.pth* to carry out the aforementioned steps.

Below is an example:

Thresholding character region and affinity score

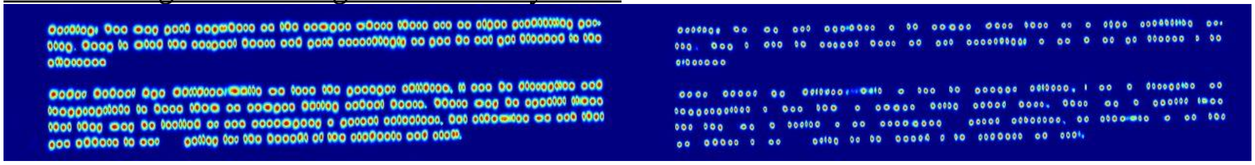
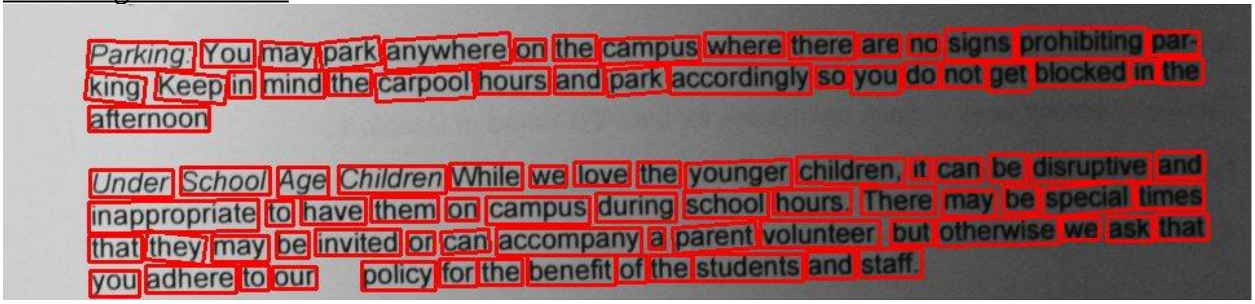


Figure 26: Thresholding character region and affinity score of Sample01

Bounding box of texts



Parking: You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

Under School Age Children While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

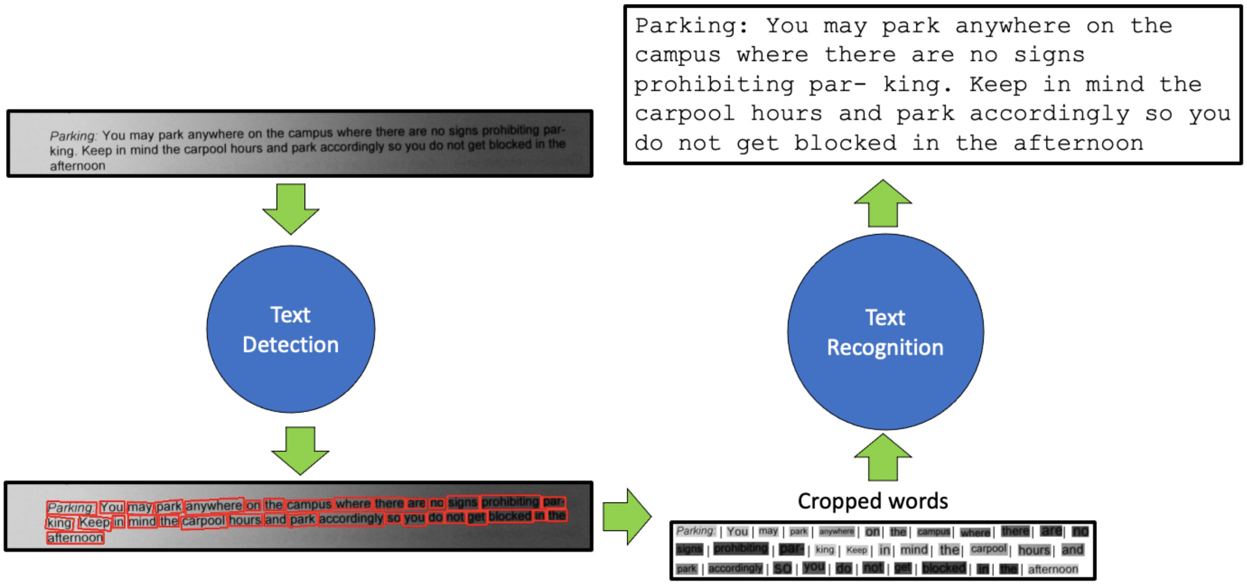
Figure 27: Image01 after bounding the words

For **text recognition**, it uses the TRBA (TPS-ResNet-BiLSTM-Attn) model. The functions of the respective sections of the model are stated below:

- **TPS** transformation: normalizes curved and perspective texts and transforms them into a standardized view
- **ResNet** feaure extrator: improved expressiveness (severe background confusion, improved in the case of a font you see for the first time)
- **BiLSTM** sequence modeling: Ignoring truncated characters that are irrelevant
- **Attn** prediction: Find missing or missing characters

We will be using another pre-trained TRBA model, *TPS-ResNet-BiLSTM-Attn-case-sensitive.pth*. The pre-trained model will predict words from the cropped image (case sensitive) and calculate each text's confidence level.

Below is a summary of the full pipeline for the CLOVA OCR algorithm:

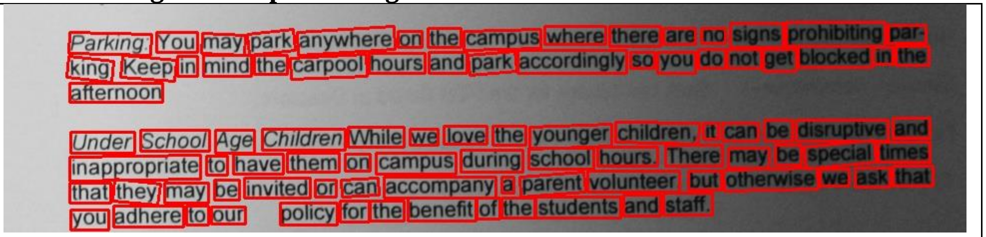
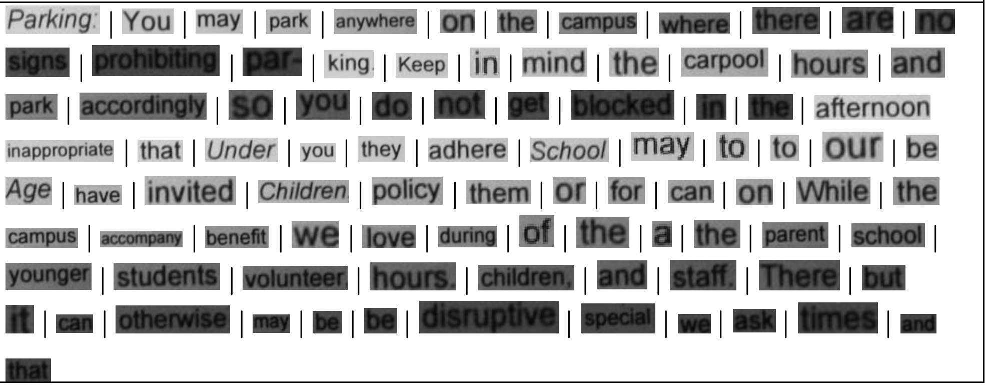


```
Parking: You may park anywhere on the
campus where there are no signs
prohibiting par- king. Keep in mind the
carpool hours and park accordingly so you
do not get blocked in the afternoon
```

Figure 28: The flow of CLOVA Algorithm

2.5.1 Sample01

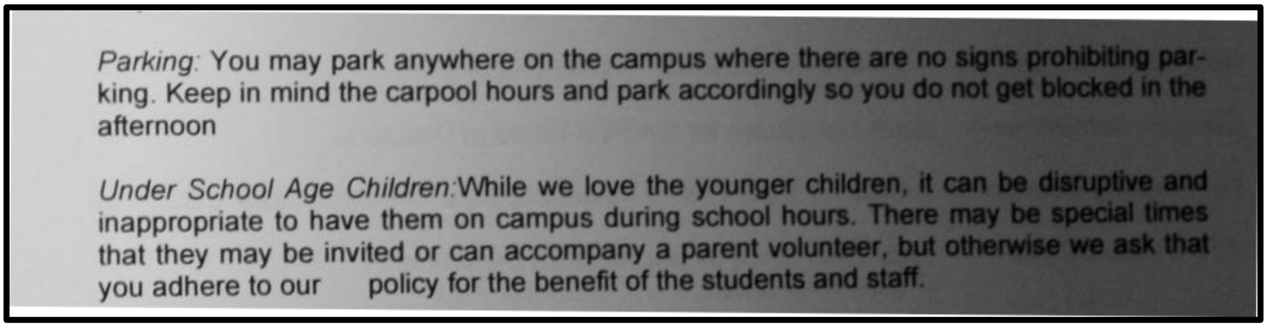
Table 16: Sample01: CLOVA – original Sample01 image

Text Bounding	 Parking: You may park anywhere on the campus where there are no signs prohibiting par- king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon Under School Age Children While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff. **Figure 29: Sample01 after bounding the words**
Cropped words	

Text predicted by TRBA model	`Parking: You may park anywhere on the campus where there are no signs prohibiting par- king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon inappropriate that Under you they adhere School may to to our be Age have invited Children. policy them or for can on While the campus accompany benefit we love during of the a the parent school younger students volunteer, hours. children, and staff. There but it can otherwise may be be disruptive special we ask times and that`

Results and observations:
For the second half of the image, the words seem to be jumbled up, even though it can accurately identify the words. We realized that the reason is that the picture is a little slanted, which disrupts the ordering of the words. Therefore, we decided to rotate the image clockwise by 0.5 degrees to ensure that the text is straight.

2.5.2 Sample01 – Rotate image clockwise by 0.5 degrees

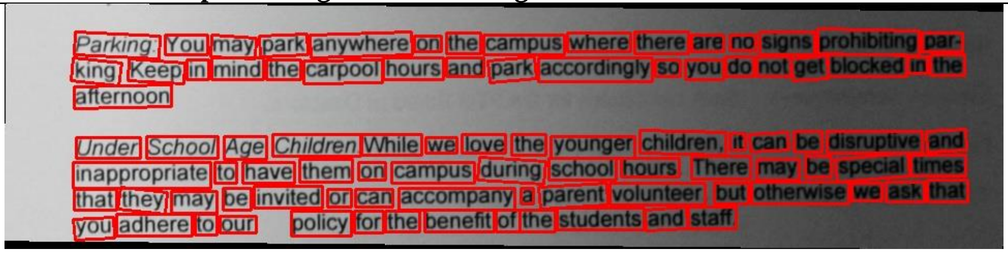
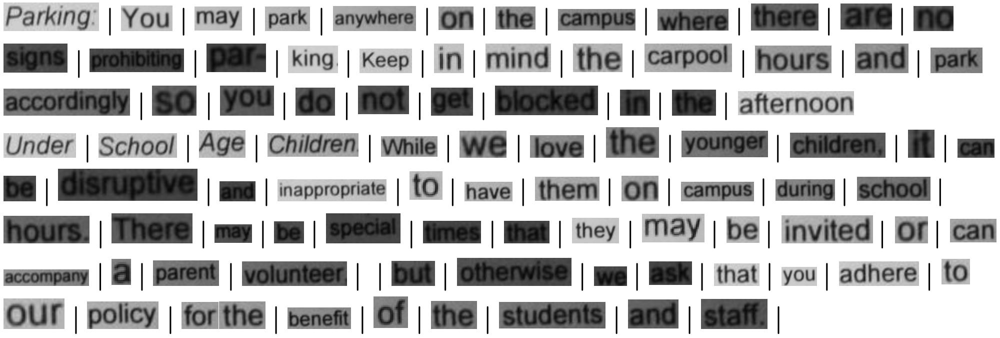


Parking: You may park anywhere on the campus where there are no signs prohibiting par- king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon

Under School Age Children: While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

Figure 30: Sample01 after rotating it by 0.5 degrees

Table 17: Sample01: CLOVA – sample01 image rotated 0.5 degrees clockwise

<table>
<tr><td>Text Bounding</td><td>
Parking: You may park anywhere on the campus where there are no signs prohibiting par- king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon
Under School Age Children While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours. There may be special times that they may be invited or can accompany a parent volunteer but otherwise we ask that you adhere to our policy for the benefit of the students and staff.
<b>Figure 31: Sample01 after bounding the words</b></td></tr>
<tr><td>Cropped words</td><td>
Parking: | You | may | park | anywhere | on | the | campus | where | there | are | no
signs | prohibiting | par- | king. | Keep | in | mind | the | carpool | hours | and | park
accordingly | so | you | do | not | get | blocked | in | the | afternoon
Under | School | Age | Children | While | we | love | the | younger | children, | it | can
be | disruptive | and | inappropriate | to | have | them | on | campus | during | school |
hours. | There | may | be | special | times | that | they | may | be | invited | or | can
accompany | a | parent | volunteer. | | but | otherwise | we | ask | that | you | adhere | to
our | policy | for the | benefit | of | the | students | and | staff. |</td></tr>
<tr><td>Text predicted by TRBA model</td><td>Parking: You may park anywhere on the campus where there are no signs prohibiting par- king. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon</td></tr>
</table>

	Under School Age Children: While we love the younger children, it can be disruptive and inappropriate to have them on campus during school hours There may be special times that they may be invited or can accompany a parent volunteer, but otherwise we ask that you adhere to our policy for the benefit of the students and staff.

Number of words correctly identified: 91 words (100%)

Results and observations:
As you can see from the results, after rotating the image, we are able to correct the order of the text. The TRBA model can determine correctly (100 percent accuracy) the words in the picture ‘sample01.png’. It also includes punctuation marks such as “:” , “.” and “,” from the text as well as case sensitivity of the text.

2.5.3 Sample02

Table 18: Sample02: CLOVA – sample02 original image

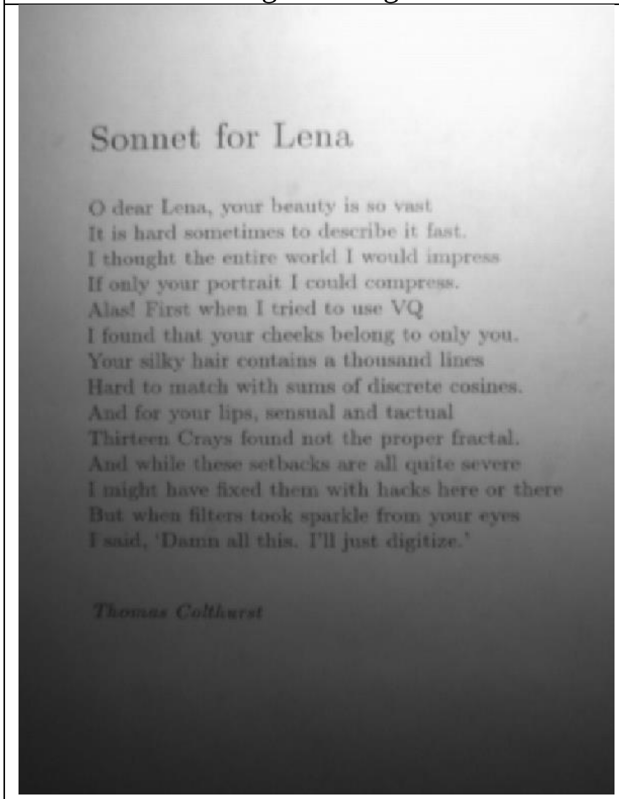
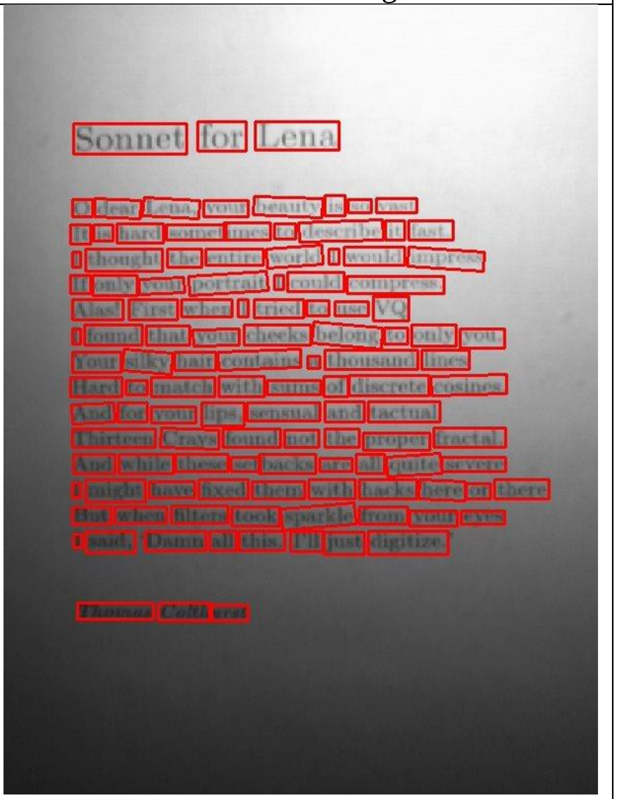
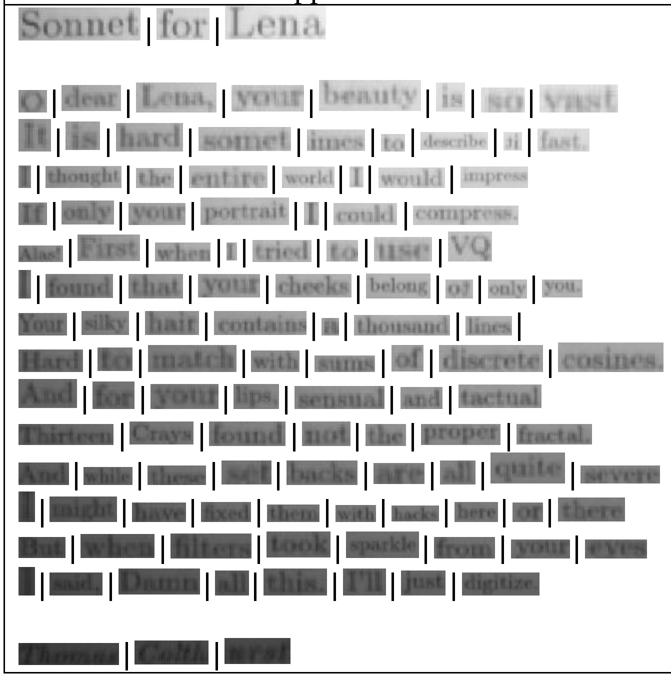
Original Image	Text Bounding
 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst	 Sonnet for Lena O dear Lena, your beauty is so vast It is hard sometimes to describe it fast. I thought the entire world I would impress If only your portrait I could compress. Alas! First when I tried to use VQ I found that your cheeks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, 'Damn all this. I'll just digitize.' Thomas Colthurst
Figure 32: Original Sample02	**Figure 33: Text bounding of Sample02**

Table 19: Sample02: CLOVA – sample02 original image

Cropped words	Text predicted by TRBA model
	Sonnet for Lena O dear Lenu, your beauly is so vast It is hard somet lines to describe it fast, I thought the entire world I would impress the only your portrait I could compross. Alas! First when I tried to use Vq I found that your checks belong to only you. Your silky hair contains a thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these set backs are all quite sover a might have fixed them with hacks here or there But when filters took sparkle from your eyes I said, Damn all this. I'll just digitize. Thomas Colth nost

Number of words correctly identified: 102 words (87.9%)

Results and observations:

As expected, sample02 performed worse than the sample01. There are more spelling errors, lack of punctuations, and some missing capital letters. This is due to the less sharp image of the sample02 image. Despite this, it still did better than the other methods used (e.g., adaptive gaussian thresholding, adaptive mean thresholding)

2.5.4 Sample02 after Cropping, CHALE Histogram Equalization and brightening

We have used CLOVA OCR in Figure 24 (Sample02 after Cropping, CHALE Histogram Equalization and brightening). Below are the results. We have also shown the text predicted by TesseractOCR here to compare between the two OCR algorithms for easier reference. As we can see CLOVA OCR performed significantly better than Tesseract OCR. This shows that CLOVA OCR can perform a lot better than Tesseract OCR as it can achieve 100 percent accuracy for the original image of Sample01 and a 94 percent accuracy for Sample02.

Table 20: Sample02: CLOVA – after Cropping, CHALE Histogram Equalization and brightening

Text predicted by TesseractOCR	Text predicted by CLOVA OCR
tSonnet for Lena	Sonnet for Lena
O dear Lena, yout heatty is so vast	o dear Lena, your beauly is so vast
	It is hard sometimes to describe it fast.
Tt in hard sometimes to describe it fast,	I thought the entire world I would impress
	If only your portrait I could comprops.
1 thought the entire world | would impress	Alas! First when I tried to use Vq
H only your portrait | could compress.	I found that your checks belong to only you.
Alas! First when | tried to use VQ	Your silky hair contains a thousand lines
	Hard to malch with sums of discrote cosines.
1 found that your cheeks belong to only you.	And for your lips, sensual and tactual

Your silky hait contains s thousand lines Hard to match with sums of discrete cosines. And for your lips, sensual and tactual Thirteen Crays found not the proper fractal. And while these sotbacks are all quite sovere 1 might have fixed them with hacks here or the re But when filters took spackle from your eyes 1 wnbl, ‘Dasun all thin. T just digitise.” Thomas Colthwrst	Thirteen Crays found not the proper fractal. And while these setbacks are all qui to soyers I might have fixed them with hacks here or there But when filters took sparkle from your eyes I said Damn all this. I'll just digitize Thomas Colthurst

Number of words correctly identified: 109 words (94%)

2.6 Summary of results

Below is the summary of the results of all the methods that we have used in this project. There are 91 words in Sample01 and 116 words in Sample02.

Table 21: Summary of results

Number of words correctly identified (Accuracy Percentage)		
	Sample01	Sample02
Tesseract OCR		
Otsu Global Thresholding	45 (49.5%)	3 (2.6%)
Adaptive Mean Thresholding	2 (2.2%)	13 (11.2%)
Adaptive Gaussian Thresholding	6 (6.6%)	16 (13.8%)
Increasing brightness	91 (100 %)	37 (32.0%)
CHALE Histogram equalization + Cropping + Increasing brightness	- did not apply it since the accuracy is already 100%	92 (79%)
CLOVA OCR		
Used original images	91 (100%)	102 (87.9%)
CHALE Histogram equalization + Cropping + Increasing brightness	- did not apply it since the accuracy is already 100%	109 (94%)

Overall, through image pre-processing and through the use of CLOVA OCR, we can have an accuracy percentage of 100% for Sample01 and 94% for Sample02. It is understandable that the accuracy for Sample02 could not be as high as Sample01 as the quality of the original image is not as good, and this makes image pre-processing of Sample01 more difficult.

3 Improvements

Discuss how to improve recognition algorithms for more robust and accurate character recognition while document images suffer from different types of image degradation.

3.1 Image pre-processing

As mentioned above, pre-processing is important in order to improve the clarity and quality of the image. We can do some pre-processing, increasing brightness, text skew correction, increasing contrast (sharpness) through unsharp masking, histogram equalization, et cetera.

3.2 Fuzzy String matching to handle spelling errors[1]

Being able to handle spelling errors is a useful way to increase the accuracy of the OCR. If we look at the text, OCR is predicted to Sample02 in sections 2.4 and 2.5.4; the accuracy is 94 percent. For the words incorrectly identified, most of the words have minimal spelling errors where one letter of the entire word is incorrectly identified. Therefore, these spelling errors can be easily handled through Fuzzy string matching. Fuzzy string matching corrects words by comparing close matching words and replacing the incorrectly spelled word with the word most suitable. It determines the closest matching words by analyzing common errors such as:

- the transposition of two letters.
- the use of incorrect vowels.
- pressing an adjacent key on the keyboard.

This is very appropriate for our case where there are minimal spelling errors where one letter of the entire word is incorrectly identified.

3.3 Using mixed CNN-LSTM network instead of only LSTM[2]

Tesseract currently uses LSTM based models. Since convolutional neural networks (CNN) showed an outstanding performance on many image processing tasks, e.g., Mane and Kulkarni (2017), training a mixed CNN-LSTM network may be able to increase the performance of OCR. Combining those diverse network structures is promising because CNNs are suited for hierarchical but location invariant tasks, whereas LSTMs are perfect at modeling temporal sequences.

We aim to train a mixed CNN-LSTM network to increase the overall performance of OCR. Combining those diverse network structures is promising because CNNs are suited for hierarchical but location invariant tasks, whereas LSTMs are perfect at modeling temporal sequences.

To further improve the predictions, we use our Python implementation of ISRI's sequence voting tool to vote on the label sequences produced by the CTC-Greedy-Decoder for each fold. This voting aims to use the faulty sequences of different models to estimate a better one. Let us assume the following predictions for only three folds in Table 3. In the first step, all sentences are aligned as far as possible.

Table 22: An example of the improvement of using a voting mechanism. Here only three folds are used to derive a result with one instead of two errors, respectively.

	Prediction
Fold 1	`An example senience with erors`
Fold 2	`A example sentence with erors`
Fold 3	`An example entence with error`
Voted	`An example sentence with erors`

All differences will be searched and counted, the character with the highest count (number of folds) is kept, and the final sequence is computed. This algorithm can be used by sample02, in which we tweak the brightness more than once with a note it has been pre-processed, so it will not be blurry. We can tweak it to get more than 250 or 300 images with different brightness constants near the optimum constant that we had already found before, which the result was 94% accuracy. This is particularly useful for the pictures that we have as the pictures have very big brightness contrasts, and the OCR will be able to view different areas of the text clearly depending on the brightness of the image itself. For example, for sample01, having a brightness of 3 may cause the text on the right side to be viewed but not the image's left side. After that, we will divide the 250 or 300 images into 5 or 6 groups that will result in 5 or 6 folds. From there, the voting algorithm will choose the folds with the most similarities.

3.4 Independent Component Analysis (ICA)[3]

ICA is a method that tries to get image information by dividing it into three main components, foreground, middle, and background. Each layer contributes to the RGB value of a pixel with its own respective constant. Here is how a pixel is made up of three layers. Here we make the three layers to be as independent as possible between one another.

$$\begin{bmatrix} x_r(p) \\ x_g(p) \\ x_b(p) \end{bmatrix} = \begin{bmatrix} r1 & r2 & r3 \\ g1 & g2 & g3 \\ b1 & b2 & b3 \end{bmatrix} \begin{bmatrix} s_1(p) \\ s_2(p) \\ s_3(p) \end{bmatrix}$$

Figure 34: Relationship of pixel’s color and independent component

p represents a pixel in the image, s1,s2,s3 represent how each layer contributes to the pixel. The respective pixel's result on the three different layers (foreground, middle, and background) will complement one another through ICA. For example, if a pixel is not clear in the foreground but is readable in the middle ground, the middle ground result will complement the result in the foreground. ICA extracts the three components (foreground, middle, and background) and integrates these components again using ORing to get a better image. To elaborate further, for example, if a pixel is blurry where the foreground, middle ground, and background of that pixel is white, black, and white, respectively, ICA will categorize the pixel as black as black is present in one of the layers (ORing). Therefore, blurry pixels that may otherwise be classified as white will be appropriately classified even though it is black. This allows the image output to be more distinct and the words to be clearer.

Below is an example of how this algorithm works.

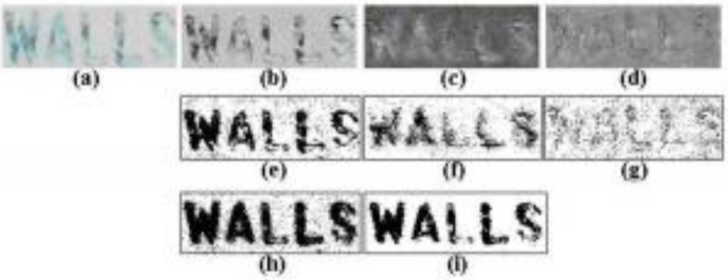


Figure 35: Proposed ICA-based processing: (a) input image, (b)-(d): foreground, middle layer, and background components obtained after execution of FastICA, (e)-(g): corresponding thresholded images, (h) integration of the foreground parts contained in the three planes, (i) smoothed image

2 issues might come while using this algorithm. First, the word might still be ambiguous, although it enhanced already. As an example

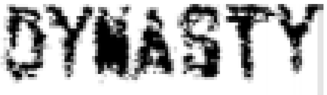


Figure 36: Example image 1

The OCR testing might read this "OYWSTY". The second reason is when the word is unavailable during word processing. For example,

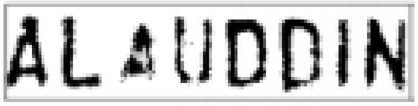
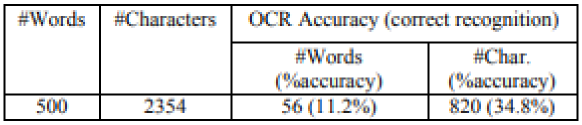


Figure 37: Example image 2

ALAUDDIN might not be recognized from our words dictionary, and it will result in the most similar word to it. The solution for this problem is to increase the domain-specific lexicon to the OCR lexicon list.

Below is the photo of OCR results tested before and after ICA.

Table 23: OCR of Camera-taken inscription images



#Words	#Characters	OCR Accuracy (correct recognition)	
		#Words (%accuracy)	#Char. (%accuracy)
500	2354	56 (11.2%)	820 (34.8%)

Table 24: OCR of images after enhancement

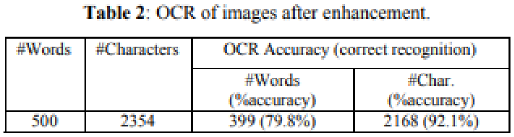


Table 2: OCR of images after enhancement.

#Words	#Characters	OCR Accuracy (correct recognition)	
		#Words (%accuracy)	#Char. (%accuracy)
500	2354	399 (79.8%)	2168 (92.1%)

The ICA can help increase the accuracy of text detection, especially in the case of sample02 where it is not 100 percent accurate. This algorithm can help us by taking more information from the middle ground and the background notes that the foreground should not as blurry as the original one when we use this algorithm.

REFERENCE

1. Hewlett Packard Enterprise Development LP. (n.d.). Fuzzy Search. Retrieved November 27, 2020,from https://www.microfocus.com/documentation/idol/IDOL/Servers/IDOLServer/11.3/Guides/html/English/expert/Content/IDOLExpert/Inquire/Fuzzy_Search.htm

2. Reul, C., Springmann, U., Wick, C., & Puppe, F. (2018). Improving OCR Accuracy on Early Printed Books by Utilizing Cross Fold Training and Voting. 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), 1-22. doi:10.1109/das.2018.30

3. Garain, U., Jain, A., Maity, A., & Chanda, B. (2008). Machine reading of camera-held low quality text images: An ICA-based image enhancement approach for improving OCR accuracy [Abstract]. *2008 19th International Conference on Pattern Recognition*. doi:10.1109/icpr.2008.4761840