

Table of Contents

1	Introduction	2
2	Methods	3
2.1	Tesseract OCR [1]	3
2.2	Levenshtein Ratio	4
2.3	OTSU [2]	5
2.4	Adaptive Mean Thresholding [3]	6
2.5	Regression Thresholding	7
3	Experiments and Results	9
3.1	Without Preprocessing	10
3.2	OTSU	11
3.3	Adaptive Mean Thresholding	12
3.4	Regression Thresholding	14
3.4.1	Linear Regression	14
3.4.2	Polynomial Regression	18
3.4.3	Neural Network	22
4	Summary and Conclusion	26
5	References	28

1 Introduction

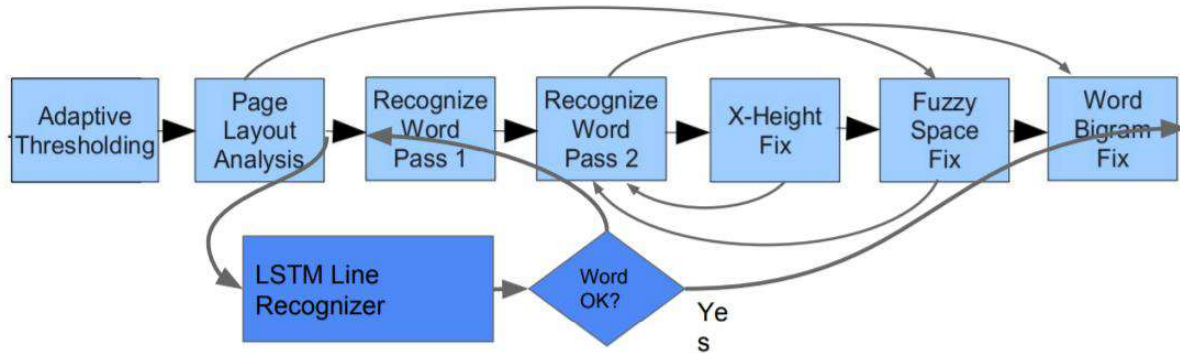
The objective of optical character recognition (OCR) is to identify and recognize texts in images. Tesseract is one of the famous open source OCR engines developed by Google, the latest version of Tesseract (version 4) can recognize text from 116 different languages.

One of the biggest challenges in OCR is to recognize text in poor quality images, such as blurry images, low contrast between text and background, text documents with shadow etc. In order to improve the effectiveness of OCR, image preprocessing steps could be carried out before feeding the image into the OCR engine.

In this project, several image binarization algorithms were explored and developed as a preprocessing step to segment poor quality image documents into black text white background image. The binarized image will then be fed into Tesseract OCR engine for text recognition. The final aim of this project is to increase OCR accuracy by preprocessing images using text segmentation. To evaluate the accuracy of OCR, Levenshtein ratio is used.

2 Methods

2.1 Tesseract OCR [1]



The image above shows a simplified text recognition pipeline of Tesseract. Tesseract can accept both grayscale and colored image, which will eventually be converted into a binary image using adaptive thresholding method.

Then, connected component analysis is carried out to identify blobs of characters and their outlines. Blobs are organized into text lines which are then analyzed to differentiate types of character spacing (fixed pitch or proportional). Fixed pitch texts are chopped into characters directly using character cells, whereas proportional text is broken into words using definite and fuzzy spaces.

Text recognition is carried out as a two-pass process. During the first pass, each word is recognized in turn and the satisfactory words are passed to an adaptive classifier. This allows the classifier to recognize text occur at the end of the page more accurately. The second pass allows words at the top of the page that are not recognized well enough in the first pass to be recognized again.

2.2 Levenshtein Ratio

Levenshtein distance is a metric that measures the similarity between 2 texts, the smaller the Levenshtein distance, the higher the similarity. It is formulated as the cost of transforming one text to the other, using only two operations (insertion and deletion), each with a cost of 1. Replacement of a character involves a deletion and an insertion, therefore bearing a cost of 2. Suppose we have 2 texts, t_a and t_b , the cost of transforming t_a to t_b or t_b to t_a is the same. Examples of determining the Levenshtein distance between 2 texts are illustrated below:

<p>Transforming “ab” to “a” requires 1 operation to delete the character “b”.</p> <p>Transforming “a” to “ab” requires 1 operation to insert the character “b”.</p> <p>Thus, the Levenshtein distance between “ab” and “a” is 1.</p>	<p>Transforming “ab” to “ac” requires 2 operations: Delete the character “b” and insert the character “c”.</p> <p>Transforming “ac” to “ab” requires 2 operations: Delete the character “c” and insert the character “b”.</p> <p>Thus, the Levenshtein distance between “ab” and “ac” is 2.</p>
--	---

The Levenshtein ratio, r between 2 texts is calculated using the formula below where $d(t_a, t_b)$ is the Levenshtein distance between texts t_a and t_b and $l(t)$ is the number of characters in the text t .

$$r = 1 - \frac{d(t_a, t_b)}{l(t_a) + l(t_b)}$$

When 2 texts are completely the same, the Levenshtein ratio is 1. When 2 texts are completely different, the Levenshtein ratio is 0.

2.3 OTSU [2]

OTSU thresholding is one of the simplest forms of image binarization methods. The threshold to separate pixels into black and white is determined by minimizing the intra-class variance or equivalently, maximizing the inter-class variance. The objective formula to find the best threshold is described below, where $\sigma_0^2(t)$ and $\sigma_1^2(t)$ are the intensity variance of class 0 and 1 separated on threshold t , and $w_0(t)$ and $w_1(t)$ are the probabilities of the 2 classes respectively.

$$\hat{t} = \min_t (w_0(t)\sigma_0^2(t) + w_1(t)\sigma_1^2(t))$$

$$w_0(t) = \sum_{i=0}^{t-1} p(i), \quad w_1(t) = \sum_{i=t}^{L-1} p(i)$$

$$\mu_0(t) = \frac{\sum_{i=0}^{t-1} ip(i)}{w_0(t)}, \quad \mu_1(t) = \frac{\sum_{i=t}^{L-1} ip(i)}{w_1(t)}$$

$$\sigma_0^2(t) = \frac{\sum_{i=0}^{t-1} (i - \mu_0(t))^2 p(i)}{w_0(t)}, \quad \sigma_1^2(t) = \frac{\sum_{i=t}^{L-1} (i - \mu_1(t))^2 p(i)}{w_1(t)}$$

After obtaining the best threshold \hat{t} , the image can be binarized using the formula below where $u(x,y)$ is the original pixel intensity at coordinate (x,y) and $f(x,y)$ is the binarized pixel intensity at pixel coordinate (x,y) .

$$f(x,y) = \begin{cases} 255 & u(x,y) \geq \hat{t} \\ 0 & u(x,y) < \hat{t} \end{cases}$$

OTSU thresholding method is a global thresholding method. It might not perform well on situations where:

- Image is noisy and the peaks of the bimodal histogram is not sharp enough.
- The object area in the image is small compared to the background area, which increases the probability of a pixel to be classified as a background.
- The illumination of the image is non-uniform.

2.4 Adaptive Mean Thresholding [3]

Adaptive thresholding method calculates threshold for each pixel by considering its neighborhood pixels. This allows the pixels at different region to have a threshold adapted to its local intensity condition and addresses the problem due to a single global threshold. The easiest implementation for adaptive thresholding is to use the mean intensity of $K \times K$ neighborhood as the threshold for a pixel.

The formula for adaptive mean thresholding is given below, where $t(x, y)$ is the threshold at pixel coordinate (x, y) and $u(i, j)$ is the pixel intensity at coordinates (i, j) .

$$t(x, y) = \frac{\sum_{i=x-\frac{K}{2}}^{x+\frac{K}{2}} \sum_{j=y-\frac{K}{2}}^{y+\frac{K}{2}} u(i, j)}{K^2}$$

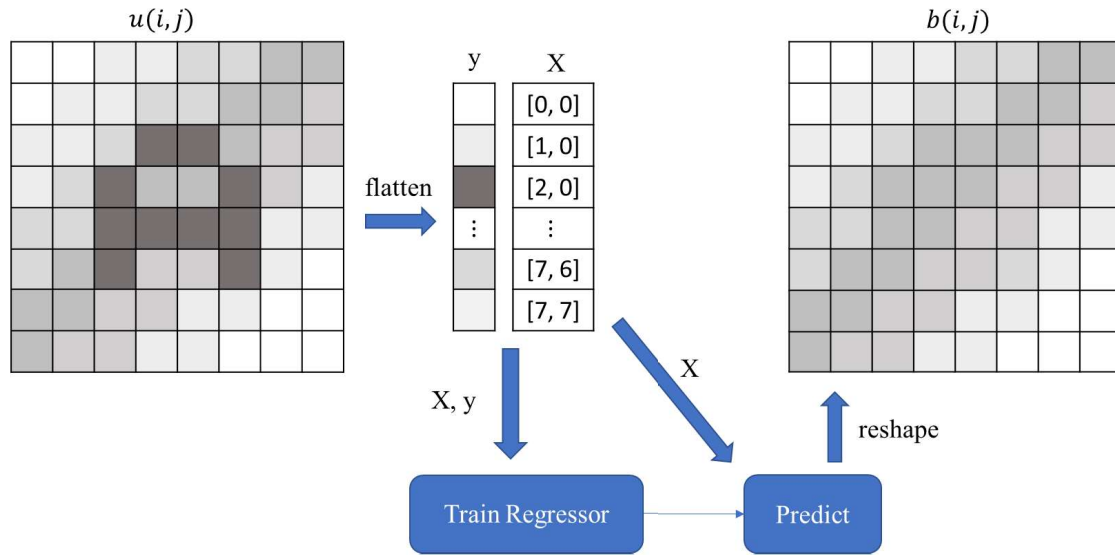
After obtaining the threshold value, $t(x, y)$ for each pixel, a constant C is subtracted from the threshold value and the image is binarized using the formula:

$$f(x, y) = \begin{cases} 255 & u(x, y) \geq t(x, y) - C \\ 0 & u(x, y) < t(x, y) - C \end{cases}$$

2.5 Regression Thresholding

The purpose of introducing regression is to predict the background intensity at each pixel.[4] This involves training a regressor such as linear regression, polynomial regression, or support vector machines, that takes in pixel coordinates as its input features and outputs the predicted background intensities at those coordinates.

To train the regression model, we use pixel coordinates as the input features, X and the original pixel intensities in the image as the target, y . The regressor's objective function is to minimize the mean square error between the predicted background intensities and the original pixel intensities.



The objective function to find the best regressor parameters $\hat{\theta}$ is expressed as the formula below, where $u(i,j)$ is the original pixel intensity at coordinate (i,j) and $b(i,j,\theta)$ is the predicted background intensity at coordinate (i,j) by the regressor with parameter θ .

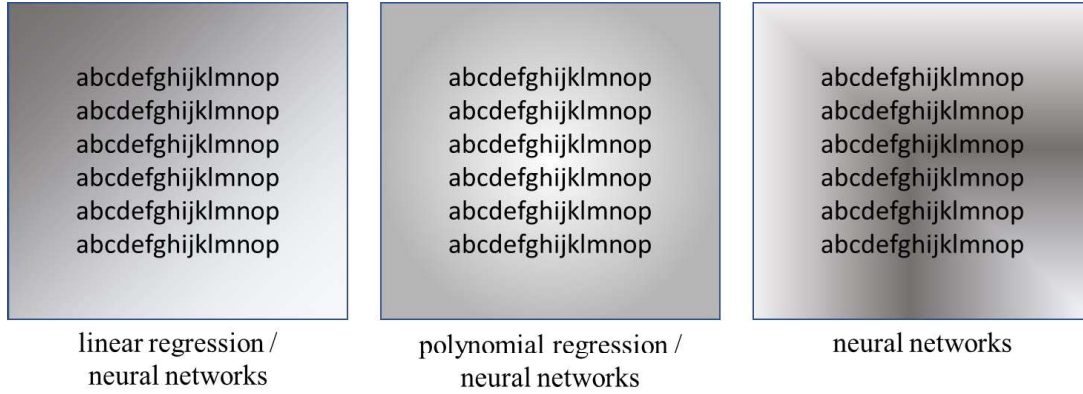
$$\hat{\theta} = \min_{\theta} \sum_x \sum_y (b(i,j,\theta) - u(i,j))^2$$

Therefore, it is important to ensure that the fraction of background pixels outweighs the fraction of text pixels in the image to allow effective modelling of the background pixel intensities over different regions of the image. The foreground (text) pixels can be viewed as noise when training the regressor. Hence, some regularization measures could be taken to prevent the regressor from overfitting to the text pixels.

After obtaining the predicted background intensities, $b(i,j,\hat{\theta})$ for each coordinate (i,j) , we could perform thresholding using the formula below, where C is a constant.

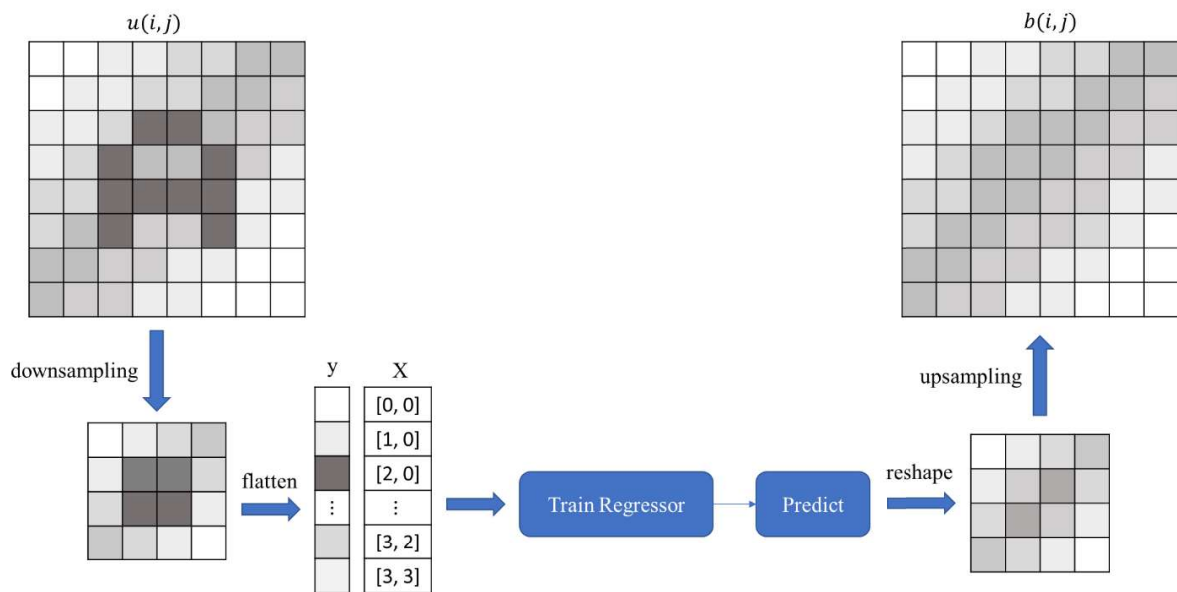
$$f(x,y) = \begin{cases} 255 & u(i,j) \geq b(i,j,\hat{\theta}) - C \\ 0 & u(i,j) < b(i,j,\hat{\theta}) - C \end{cases}$$

The advantage of this method is that different regressors could be used for different background complexity. For instance, linear regression could be used for linearly changing background intensities over a constant direction; polynomial regression could be used for radial background intensity gradient; neural networks could be used for more complex backgrounds.



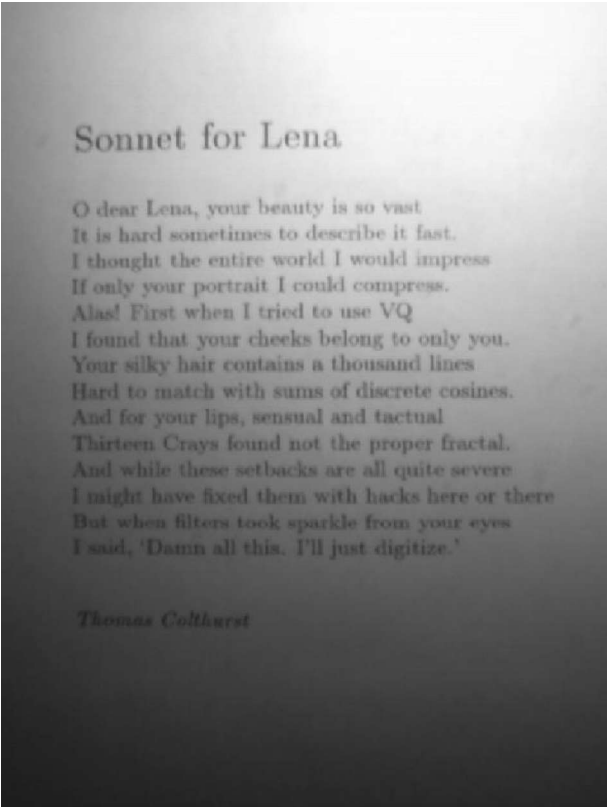
However, a new regression model has to be trained for every image. This causes problem for some models such as neural networks, ensemble or boosting regressors which take a long time to train. Besides, some regression models might require plenty of hyperparameter tuning in order to achieve a satisfactory segmentation result. Hence, it is important to choose the most suitable regression model for this method. In this project, the performances of linear regression, polynomial regression and neural networks for different image background complexity are compared.

Image could be downsampled or pooled so that less pixels are being used for training, which helps in reducing training time and improving model generalization. The tradeoff associated with this is the loss of resolution. Upsampling has to be carried out at the final stage to obtain a threshold map with the same size as the original image. Bilinear interpolation could be used during the upsampling stage to better approximate the intensities of missing pixels.



3 Experiments and Results

In the following experiments, we will be using the 2 image samples shown below.

<i>Sample01</i>	<p><i>Parking:</i> 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</p> <p><i>Under School Age Children:</i> 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.</p>
<i>Sample02</i>	 <p>Sonnet for Lena</p> <p>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 tactical 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.'</p> <p>Thomas Colthurst</p>

3.1 Without Preprocessing

Firstly, the sample images are fed directly to the Tesseract OCR engine without any preprocessing. From the result, we can see that Tesseract fails to detect almost all text in *Sample02* which is blurry. Whereas, for *Sample01*, Tesseract manage to detect text in regions where the contrast between background and text are still significant, but for regions with darker background, Tesseract is unable to detect text within it.

Sample	Text Detected	Levenshtein Ratio
<i>Sample01</i>	INFO: Parking: You may park anywhere on the ce king. Keep in mind the carpool hours and park 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 of	0.66242
<i>Sample02</i>	Sonnet for Lena	0.05247

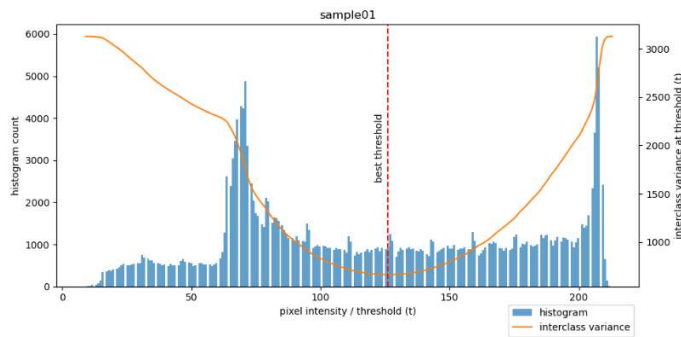
3.2 OTSU

OTSU algorithm could not provide satisfactory text segmentation results for all sample images due to the non-uniform illumination. Therefore, the text recognized by Tesseract is also not accurate enough and there is no improvement as compared with not performing preprocessing at all. The segmentation result and text recognized for each sample image is shown below.

Sample01

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
you adhere to our policy for the benefit of



Levenshtein Ratio: 0.65478

Text Detected:

Parking You may park anywhere on
the cf
king. Keep in mind the carpool hours
and peri,
afternoon

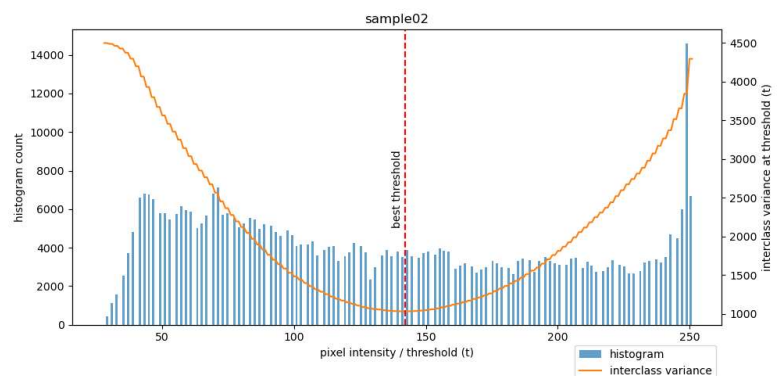
Under School Age Children: While we
love
inappropriate to have them on campus
@ i
that they may be invited or can
accompany J
you adhere to our policy for the
benefit of

Sample02



Levenshtein Ratio: 0.04946

Text Detected: Sonnet for ler



3.3 Adaptive Mean Thresholding

After using adaptive mean thresholding, the text segmentation results improved significantly. This is due to the adaptive threshold which is calculated based on the local region properties of the pixels. Therefore, it is not affected much by the non-uniform illumination problem and is able to differentiate between background and text pixels more effectively.

Sample01

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.

Levenshtein Ratio:
0.97987

Text Detected:

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 : : a

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: .

Sample02

Levenshtein Ratio: 0.85692

Sonnet for Lena

O dear Lena, your beauty is in vail
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 Culbertson

Text Detected:

Sonnet for Lena

O dear Lenn, your benuly iat ais vol
Tels hard sometiines to describe tt fast.
Lthought the entice woekl T weukl improas
If ony your portrait [caukd compress.

. Alas! First when T tried to use VQ
I found that your cheeks Leloug te only you,
Your silky hair couloing nv thousand [ines
Ilan to imatel with sums of diserete cosines. .
And for your Ips, sensual and (netual
Vairtven Crays found aot the proper Gactal,
Aud white these setbacks are all cnite severe
Lialght bave dixed them with hacks here or there
Bot when filters took sparkle from your exes
Tented, Dainn all this. TMU just digitize.◆

However, we can see tiny noises appear on the binarized images. To reduce the noise, Gaussian filter can be applied to the image prior to computing the adaptive mean threshold. The results below illustrates the improvement after using Gaussian filter.

Sample01

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.

Levenshtein Ratio:
0.98742

Text Detected:

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.

Sample02

Levenshtein Ratio: 0.92806

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

Text Detected:

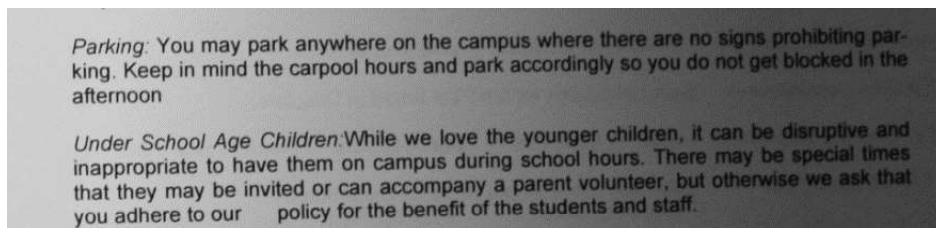
Sonnet for Lena

O dear Lena, your beauty in so vast
It is hare sametinaes to cdescribe ft fant.
Ethought the entire world f would impress
If only your portrait [could compress,
Alas! First when [tried to nse VQ
1 found that your cheeks belong tv only you.
Your silky hair contains n thonsantd fines
Ifarl to match with suing of diserete cosines,
And for your lips, sensual and dactual
Thirteen Crays found nol the proper fractal,
And while these setbacks are all quite severe
Tanight lave fixed them with lineks here or there
But when fillers took sparkle from your eyes
Tsaid, Damn all this. UL just digitize.❖

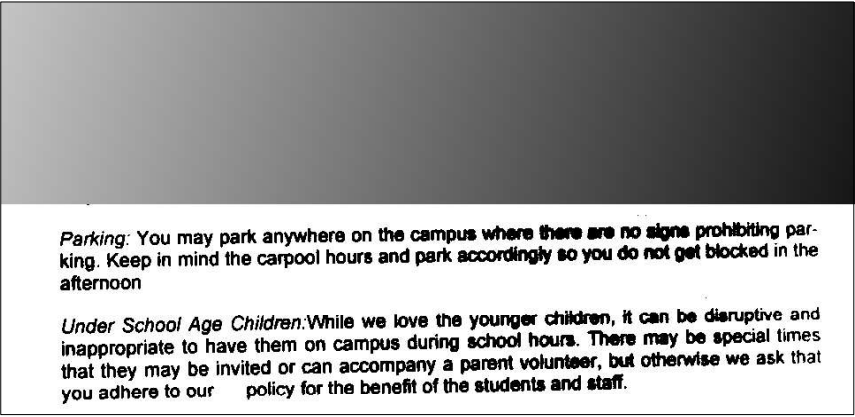
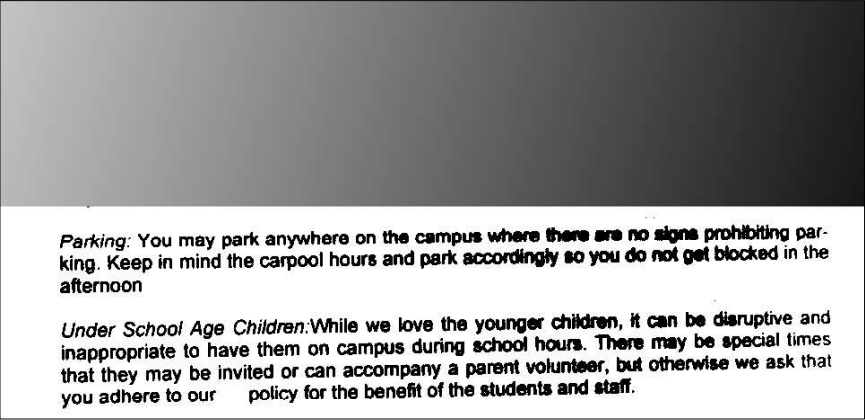
3.4 Regression Thresholding

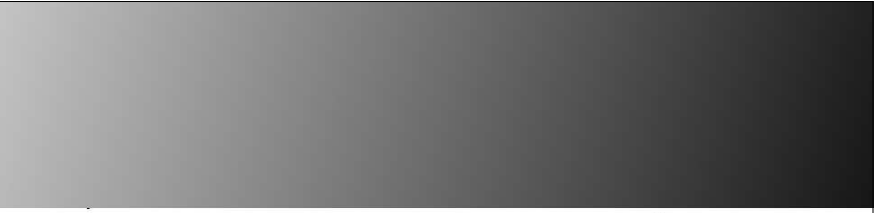

3.4.1 Linear Regression

Linear regression could be used to estimate the background intensity for linearly changing background gradient over a constant direction. *Sample01* as shown in the image below is an example that demonstrates this kind of gradient. The background gradient is linearly changing from higher intensity to lower intensity over a constant direction from left to right.







The background estimated using linear regression and the binarized image on *Sample01* with different downsampling factor are shown below:

Downsampling Factor	Background Estimated / Binarized Image
1 (Without Downsampling)	
2	

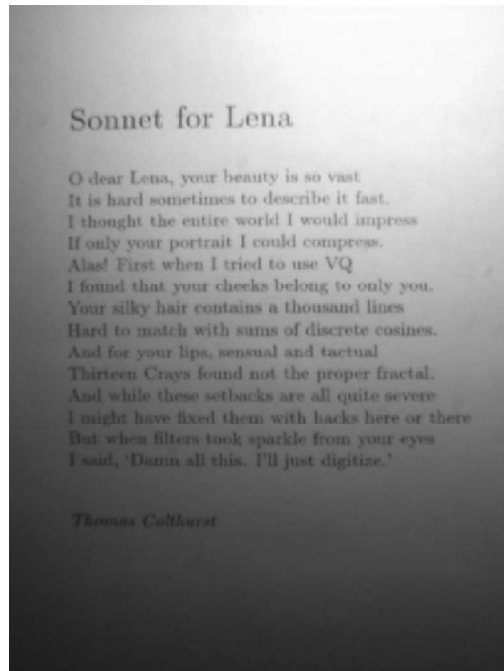
5	 <p><i>Parking:</i> You may park anywhere on the campus where there are no signs prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon</p> <p><i>Under School Age Children:</i> While we love the younger children, it can be disruptive and 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.</p>
10	 <p><i>Parking:</i> 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</p> <p><i>Under School Age Children:</i> 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.</p>

From the results above, we can see that downsampling the image have minimal effect on linear regression. This is due to the use of bilinear interpolation when upsampling the background intensity map. The text detected and their Levenshtein ratio is shown in the table below:

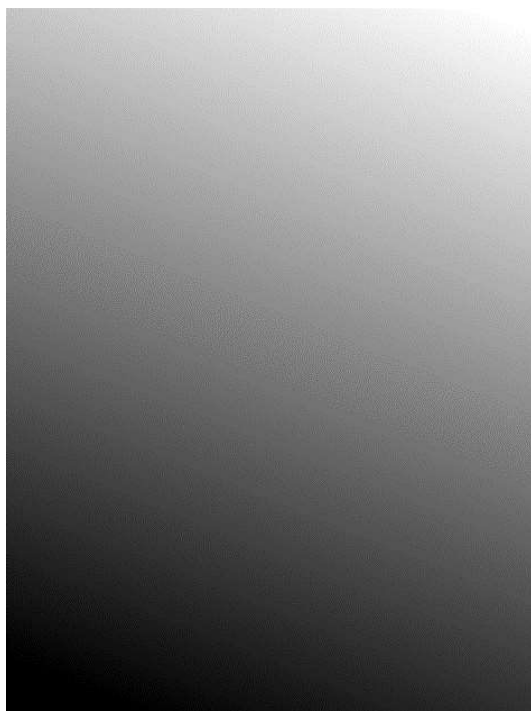
Downsampling Factor	Text Detected	Levenshtein Ratio
1	<p>Parking: You may park anywhere on the campus where there sre no signe prohibiting parking. Keep in mind the carpool hours and park accordingly so you do not get blocked in the afternoon</p> <p>Under Schoo! Age Children.While we love the younger children, it can be disruptive and inappropriate to have them on campus during schoo! 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.</p>	0.98164

2	<p>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</p> <p>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.</p>	0.98164
5	<p>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</p> <p>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.</p>	0.98357
10	<p>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</p> <p>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.</p>	0.98164

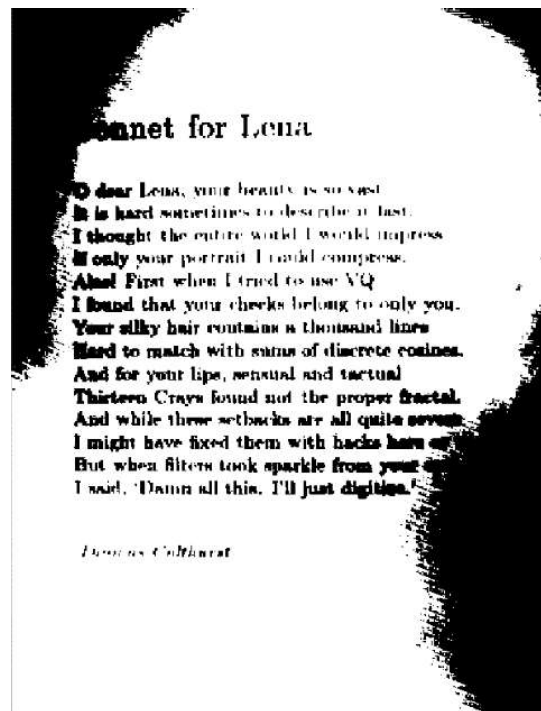
However, when linear regression is used for *Sample02*, the segmentation result is not satisfactory. This is due to the non-linear background surface property of *Sample02*. From the original image, we can observe some radial property in the background gradient. An example is shown below:



Original Image



Estimated Background

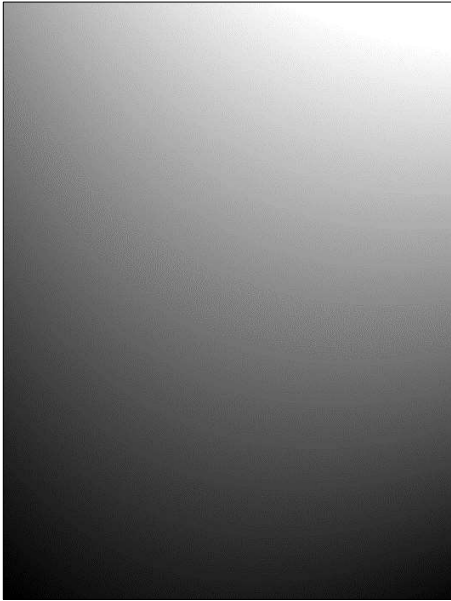
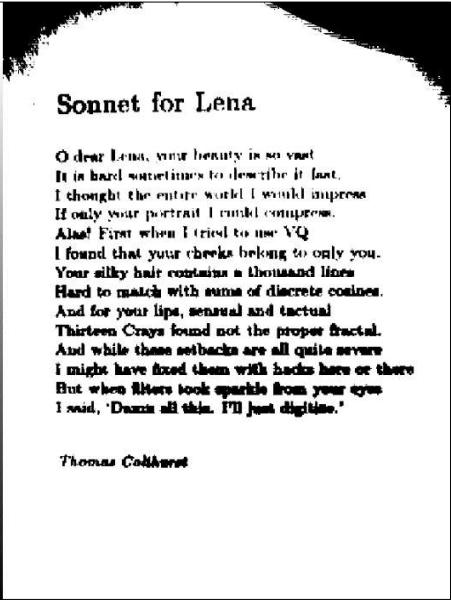

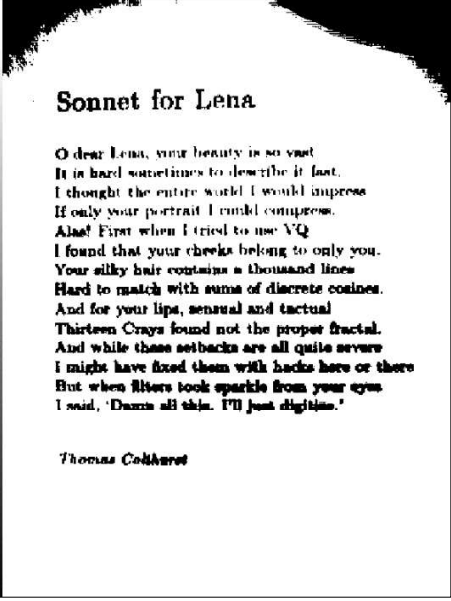


Binarized Image

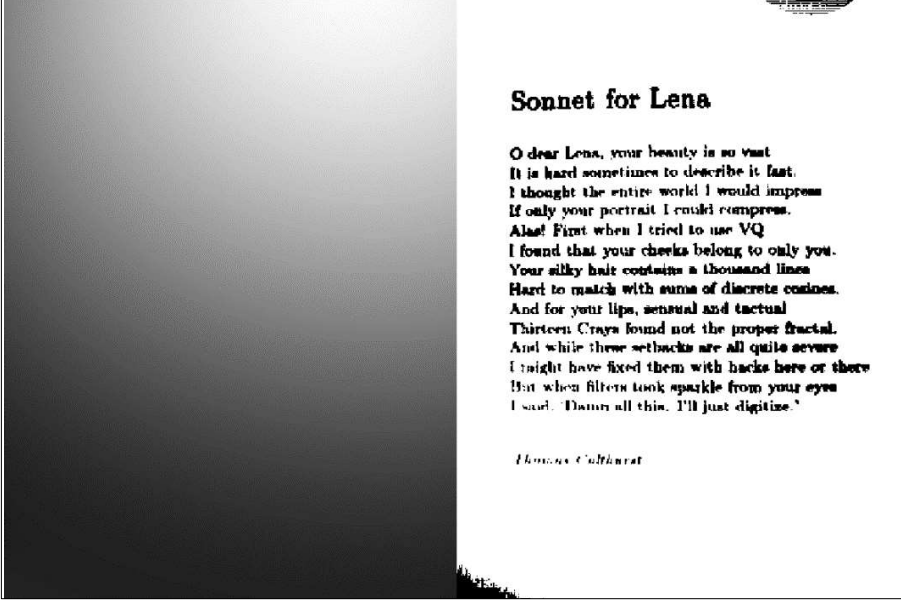
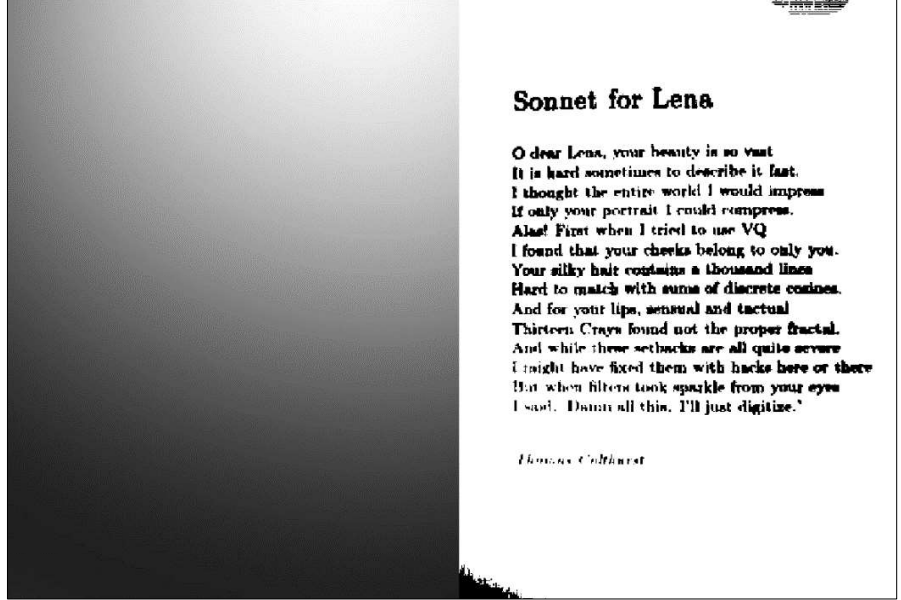
3.4.2 Polynomial Regression

As linear regression failed to model the background of *Sample02* in which its gradient exhibits radial property, polynomial regression is experimented. The results of estimated background and binarized image on *Sample02* with different downsampling factor and polynomial degree is shown below:

Polynomial Degree = 2

Downsampling Factor	Background Estimated / Binarized Image	
1		
10		

Polynomial Degree = 3

Downsampling Factor	Background Estimated / Binarized Image
1	 <p>Sonnet for Lena</p> <p>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."</p> <p><i>Thomas Culbertson</i></p>
10	 <p>Sonnet for Lena</p> <p>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."</p> <p><i>Thomas Culbertson</i></p>

From the results above, we can see that polynomial regression is also robust to downsampling, whereby the binarized image and the estimated background surface is almost identical before and after downsampling. The segmentation results for polynomial degree of 3 is better than polynomial degree of 2. The OCR results and their Levenshtein ratio are shown in the table below:

Polynomial Degree = 2

Downsampling Factor	Text Detected	Levenshtein Ratio
1	<p>Sonnet for Lena</p> <p>O dear Lena, vine beauty in se vant</p> <p>Bein bare wanetitacs to lesetibe it fant, Lehongbt the entire world [wrk impress Tf oaly your portrait cunkl compress.</p> <p>Alaa! Firat when TE (ried te nse VQ</p> <p>L found that your cheeks belong to only you. Your silky bait copteips ♦ thousnd lines Hard to malch with wura of discrete cosines. And for your lips, sensual and tactuad Thirteen Crays found not the proper fractal. And while thase setbacks are all quile severe T might have fixed them with hacks here ot there But when fers book epariie from your eye J anid, ♦Deze obi thle. FT poet digitine.♦</p> <p>Thomas Coftheret</p>	0.85261
10	<p>Sonnet for Lena</p> <p>O dear Lena, sii beauty in so vast</p> <p>Te ia bard sanetitiaes to lesetibe jt fart, Tehongbt the entre orld [wnukl anproas Vf only sour portrait cankd compress.</p> <p>Alas! Firat when [tried te ne VQ</p> <p>L found that your cheeks belong to only vou. Your silky bait contains ♦ thousand lines Hard to match with ume of diacrete conines. And for your lips, sensual and tactuad Thirteen Crays found not the proper fractal. And while thase setbacks are all quile severe E might have fixed them with hacks here or thers Blut when fiers took sparkle from your eyes T anid, ♦Demin oli thle. F♦T feet digitioe.♦</p> <p>Thomas Coftharet</p>	0.87046

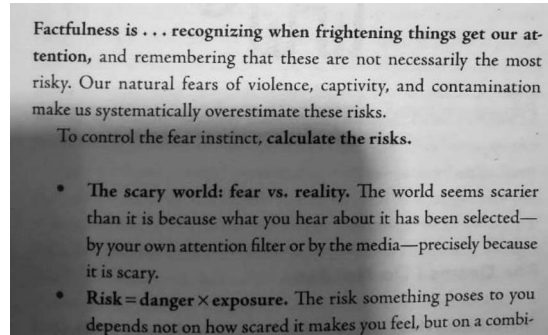
Polynomial Degree = 3

Downsampling Factor	Text Detected	Levenshtein Ratio
1	<p>Sonnet for Lena</p> <p>◆ dear Lena, your hewuty ia eo vest</p> <p>fh is hard sometimes to describe it fast,</p> <p>E thought the entice worl 1 would impress Lf only your portrait [couki romprese,</p> <p>Alest Firet when 1] tried to uae VQ</p> <p>L fowod that your cheeks belong to only you. Your sitky bait couteites @ thoussod linca Hard to match with wuma of discrete cosines. And for your lips, sensual and tectual Thirteen Craya found not the proper fractal. Ane while theee setbacks are all quile severe: Craight have fixed them with hacke bere or there Bin wher filtets took aparkle from your eye Paid. Dann all thin, VI just digitize.◆</p> <p>Phowae Culthnrat</p>	0.87421
10	<p>Sonnet for Lena</p> <p>O dear Lena, your hewuty ia eo vest</p> <p>fy is hard scinetimes to describe it fant. Etought the entice worl 1 would impress Lf only your portrait [could remprese.</p> <p>Alast Firet when I tried! to use VQ</p> <p>L fomnd that your cheeks belong to only you. Your silky bait couteins a thousand linca Hard to match with sume of discrete comines. And for your lips, sensual and tectual Thirtern Crays fowod oot the proper fractal. And while these sctacks are all quile severe: Timight have fied them with hacke bere or there Bin when Bltets took sparkle from your eyew Tsant. Dstt all thin, TU just digitize.◆</p> <p>fio Coltharet</p> <p>i</p>	0.87905

3.4.3 Neural Network

Linear regression and polynomial regression would be sufficient to model simple background. But when a more complex background is encountered, both these models might not be able to estimate background intensities accurately. An example of an image with a slightly complex background is shown below:

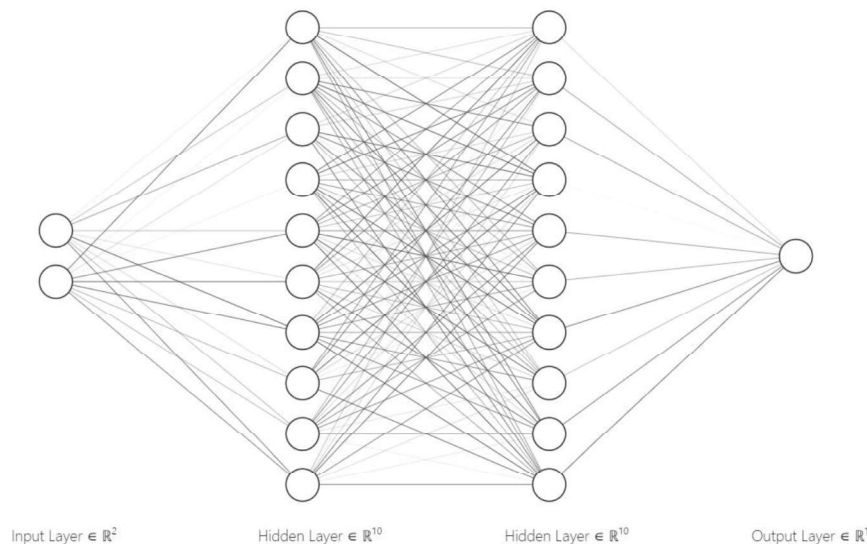
Sample03



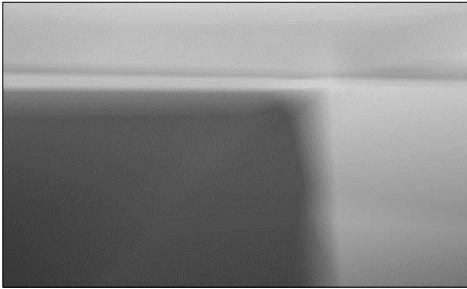
The estimated background and binarized image using linear regression and polynomial regression on *Sample03* are shown below:

Polynomial Degree	Background Estimated / Binarized Image	Levenshtein Ratio
1 (Linear)		0.73092
2		0.77461
3		0.79861

Neural networks is often a good choice for modelling data samples with high complexity. In the following experiment, a simple neural network is constructed, with 2 hidden layers and 10 nodes in each hidden layer. The input to the neural network is a 2D pixel coordinate and the output is the background intensity at the coordinate.

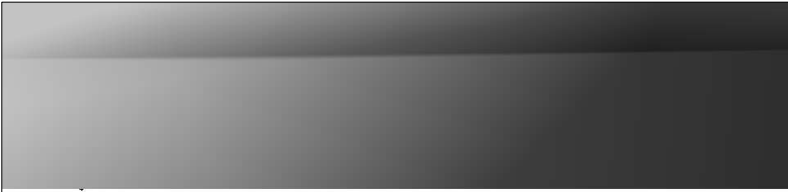


The estimated background and binarized image using the constructed neural network on *Sample03* are shown below, together with the OCR results and Levenshtein ratio.

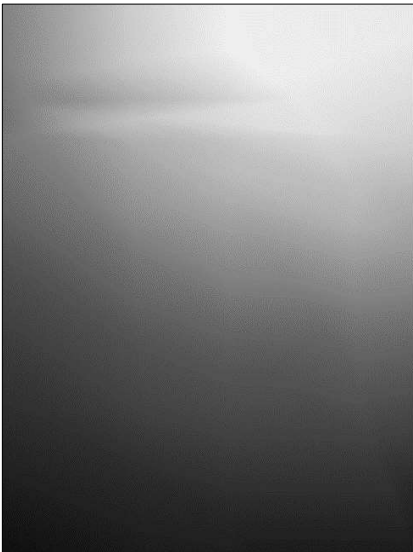
Background Estimated / Binarized Image	 <p>Factfulness is . . . recognizing when frightening things get our attention, and remembering that these are not necessarily the most risky. Our natural fears of violence, captivity, and contamination make us systematically overestimate these risks.</p> <p>To control the fear instinct, calculate the risks.</p> <ul style="list-style-type: none"> • The scary world: fear vs. reality. The world seems scarier than it is because what you hear about it has been selected—by your own attention filter or by the media—precisely because it is scary. • Risk = danger × exposure. The risk something poses to you depends not on how scared it makes you feel, but on a combi-
Text Detected	<p>Factfulness is . . . recognizing when frightening things get our attention, and remembering that these are not necessarily the most risky. Our natural fears of violence, captivity, and contamination make us systematically overestimate these risks. %</p> <p>To control the fear instinct, calculate the risks. .</p> <p>* ♦ The scary world: fear vs. reality. The world seems scarier than it is because what you hear about it has been selected♦ by your own attention filter or by the media♦precisely because</p> <p>it is scary.</p> <p>* Risk = danger x exposure. The risk something poses to you</p> <p>depends not on how scared it makes you feel, but on 2 combi-</p>
Levenshtein Ratio	0.98319

From the results above, we can see that by using neural network, we can obtain a good estimate of the background pixel intensities for more complex background patterns. Besides, neural network could also be used for simpler background estimation which is shown below using *Sample01* and *Sample02*.

Sample01

Background Estimated / Binarized Image	 <i>Parking:</i> 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 <i>Under School Age Children:</i> 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.
Text Detected	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.
Levenshtein Ratio	0.98649

Sample02

Background Estimated / Binarized 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 tectonic Thirteen Crays found not the proper fractal. And while these setbacks are all quite severe I might have fixed them with locks here or there But when filters took sparkle from your eyes I said, "Damn all this. I'll just digitize." Thomas Collyer
---	---

Text Detected	<p>Sonnet for Lena</p> <p>◆ dear Lena, your hesuty is so veat</p> <p>I in bard sometimes to ceacrite it East.</p> <p>LT thonght Ube entire world 1 would impress Lf only ◆eur portrait [couki remprees.</p> <p>Alas◆ Fired when 1 triel to use VQ</p> <p>Pfound that your checks belong to only you. Your silky hair coviaips a thonssand lines Hand to match with sutras of discrete cosines. Aud for your Lipa, sensual atid tactia) Thirteen Crave feud not the proper fractal. Ard while these setbacks aie ad] quite severe Crpight have fisel Ue with leks bere of there Bit wher Alters took sparkle from sone even T nail, (Darin all thin, 1k jurt aligitize *</p> <p>Thitnes Cotthurst</p>
Levenshtein Ratio	0.84767

However, training neural networks might take a long time and neural networks are often prone to overfitting. Therefore, in the experiments above, I have set a total number of 10 iterations, to shorten the training time and to ensure that it is sufficient enough to achieve a desirable but not overfitted background estimation.

Another thing is that neural networks are not robust to downsampling, this is because neural networks often requires large amount of data for training in order to be able to learn the background information well. The amount of pixel data used for training decreases by the square of downsample factor, therefore, it is important to ensure that the downsample factor is not too large to prevent underfitting of neural networks.

4 Summary and Conclusion

The tables below summarize the Levenshtein ratio between the detected text using Tesseract and the ground truth text after using different image binarization algorithms. The bold Levenshtein ratio is the best Levenshtein ratio obtained for each sample image.

Sample01

Image Binarization Algorithm	Downsampling Factor			
	1	2	5	10
None	0.66242	-	-	-
OTSU	0.65478	-	-	-
Adaptive Mean Thresholding	0.97987	-	-	-
Gaussian Filter + Adaptive Mean Thresholding	0.98742	-	-	-
Linear Regression (P=1)	0.98164	0.98164	0.98357	0.98164
Polynomial Regression (P=2)	0.98263	0.98269	0.98361	0.98069
Polynomial Regression (P=3)	0.98551	0.98548	0.98456	0.98357
Neural Network	0.98649	0.98842	0.98164	0.79043

Sample02

Image Binarization Algorithm	Downsampling Factor			
	1	2	5	10
None	0.05247	-	-	-
OTSU	0.04946	-	-	-
Adaptive Mean Thresholding	0.85692	-	-	-
Gaussian Filter + Adaptive Mean Thresholding	0.92806	-	-	-
Linear Regression (P=1)	0.72922	0.71001	0.76746	0.75990
Polynomial Regression (P=2)	0.85261	0.84579	0.86122	0.87046
Polynomial Regression (P=3)	0.87421	0.90119	0.87717	0.87905
Neural Network	0.84767	0.82137	0.86364	0.81535

From the tables above, in general, Gaussian filter + adaptive mean thresholding yields a consistently good image binarization result for OCR purposes.

The tables below summarize the time taken in seconds to perform various image binarization algorithms.

Sample01 – (965 × 229 = 220985 pixels)

Image Binarization Algorithm	Downsampling Factor			
	1	2	5	10
OTSU	0.01795 s	-	-	-
Adaptive Mean Thresholding	0.00203 s	-	-	-
Linear Regression (P=1)	0.07380 s	0.01596 s	0.00396 s	0.00296 s
Polynomial Regression (P=2)	0.08374 s	0.02992 s	0.00598 s	0.00499 s
Polynomial Regression (P=3)	0.12666 s	0.02893 s	0.00598 s	0.00499 s
Neural Network	5.78956 s	1.04672 s	0.16855 s	0.06283 s

Sample02 – (589 × 782 = 460598 pixels)

Image Binarization Algorithm	Downsampling Factor			
	1	2	5	10
OTSU	0.05247 s	-	-	-
Adaptive Mean Thresholding	0.00598 s	-	-	-
Linear Regression (P=1)	0.20445 s	0.03092 s	0.00798 s	0.00698 s
Polynomial Regression (P=2)	0.23650 s	0.04385 s	0.01197 s	0.00798 s
Polynomial Regression (P=3)	0.26027 s	0.06782 s	0.01895 s	0.00895 s
Neural Network	11.75618 s	2.81928 s	0.43907 s	0.12290 s

From the tables above, the fastest image binarization algorithm is adaptive mean thresholding. Therefore, we can conclude that adaptive mean thresholding is the most robust image binarization algorithm out of all being experimented as it yields consistently good results within the shortest time.

5 References

- [1] R. Smith, "An overview of the Tesseract OCR engine," in *Ninth international conference on document analysis and recognition (ICDAR 2007)*, 2007, vol. 2: IEEE, pp. 629-633.
- [2] J. Yousefi, "Image binarization using otsu thresholding algorithm," *Ontario, Canada: University of Guelph*, 2011.
- [3] T. R. Singh, S. Roy, O. I. Singh, T. Sinam, and K. Singh, "A new local adaptive thresholding technique in binarization," *arXiv preprint arXiv:1201.5227*, 2012.
- [4] G. D. Vo and C. Park, "Robust regression for image binarization under heavy noise and nonuniform background," *Pattern Recognition*, vol. 81, pp. 224-239, 2018.