

# UNIK4690 Project

Text Recognition

Akhsarbek Gozoev - akhsarbg

Sadegh Hoseinpoor - sadeghh

Key Lung Wong - keylw

Report



# Contents

<b>1</b>	<b>Introduction</b>	<b>2</b>
1.1	Project Description . . . . .	2
1.2	Project Limitations . . . . .	2
1.3	Project Components . . . . .	2
1.4	Report Layout . . . . .	3
<b>2</b>	<b>Method</b>	<b>4</b>
2.1	Description . . . . .	4
2.2	Component: Text segmentation . . . . .	4
2.3	Component: Preprocessing . . . . .	5
2.4	Component: Classification . . . . .	7
2.5	Component: Datasets . . . . .	8
<b>3</b>	<b>Discarded Method</b>	<b>10</b>
3.1	Description . . . . .	10
3.2	Text Segmentation . . . . .	10
3.3	Preprocessing . . . . .	11
3.4	Classification . . . . .	13
3.5	Datasets . . . . .	13
<b>4</b>	<b>Result - Discussion</b>	<b>15</b>
4.1	Result: Text Segmentation . . . . .	15
4.2	Result: Preprocessing . . . . .	15
4.3	Result: Classification . . . . .	17
4.4	Final Result . . . . .	18
<b>5</b>	<b>Conclusion</b>	<b>20</b>
5.1	Time pressure . . . . .	20
5.2	Ambition vs reality . . . . .	20
5.3	Final remarks . . . . .	20
<b>6</b>	<b>References</b>	<b>21</b>
6.1	Note . . . . .	21

# Chapter 1

## Introduction

### 1.1 Project Description

The purpose of our project is to recognize text from image, and be able to do string manipulation on it. Hence this falls under the “Optical character recognition” (OCR) problem

### 1.2 Project Limitations

As OCRs are still a challenging task even for companies like Google, ref. reader to Googles OCR translator application on smartphones; “Translate”, drawbacks such as; difficulty finding all the text on the photo because of lighting, noise etc., therefore we take it for granted that we should limit our text recognition problem.

#### Initial limitations

- English alphabet [upper and lower case] + numbers [0-9] + space
- Homogeneous background, white
- Skew free text
- Computer printed text
- Even lighting

### 1.3 Project Components

We have come to the conclusion that the OCR software has 3 main parts. Each part is essential, for the OCR software to be able to perform its purpose.

1. *Text segmentation*
  - Finding text segments on an image and returning the text segments
2. *Preprocessing*
  - Do preprocessing on the segmented text, such as rotation and line and symbol segmentation. Preprocessing from definition, should be done first, however because of simplification we assume we manage to segment out the text first.

### 3. Classification

- Classification of the symbols

Additionally there is one more very important component for this OCR software to work, ***labeled data***. Even though one might not need to code for this part, a good pool of labeled data is needed to be able to classify symbols. More on this in section 2.5

4. Data classification - *Gathering labeled data to train a classification algorithm*

## 1.4 Report Layout

This report is divided into 6 sections. In Chapter 1, we introduce the project and how the report is organized. Chapter 2 is where we present methods we have used in our final result. Discarded methods tried will be covered in Chapter 3. In Chapter 4 we review our approach and its results, we present the confidence in our solution and what vulnerability exist. In Chapter 5 we present our thoughts on the project; our ambition vs the actual result, and potential future improvements on the software. Chapter 6 includes reference to third party material we have used, such as; code, articles and datasets.

# Chapter 2

## Method

### 2.1 Description

The following section, as mentioned in Section 1.4, will present our methods. Below we present each component individually and the corresponding methods. A short description on what challenges have to be solved for each component will be addressed first.

### 2.2 Component: Text segmentation

#### Description

In this part of the program, we want to be able to segment out text segments in the image. Afterwards we want to segment out lines and then letters for further classification, see section 2.3.2 and 2.3.3 for reference. In this part we assume that the images have black text on white background, and it mostly consists of text.

#### Approach: Simple Image Analysis Techniques

Our approach here is inspired by this online blog [Source\[4\]](#). We simplified the original approach to the following steps:

1. **Find Edges/Outliners of the image.** Initial idea is to use Canny, but we found Morphological Gradient to perform better. It indicates the contrast of the edges, so we can get better differences in some natural images (granted the text and background are close to black and white respectively).
2. **Otsu Thresholding.** We need our image to be a binary image. We simply use OpenCV Otsu algorithm to achieve it that.
3. **Morphological Closing.** Since we want line segments we use Morphological Closing with large horizontal filter to merge as many horizontal letters together as possible.
4. **Extract Regions.** OpenCv FindContours was used to find the different text regions. We then exclude regions which are smaller than a selected threshold. The different region are returned as coordinates of the different rectangular boxes.

## 2.3 Component: Preprocessing

### Description

Definition of preprocessing; the act of preparing the data for further use, in our case for classification.

After the text segmentation we assume we have an image consisting of white text on black background. What remains for us to do is to segment out each character and format them to the right data type for the classifier. Hence we have 4 problems to solve:

### Problems

- Rotated text
  - Our approach for character segmentation needs text rotated horizontally.
- Line segmentation
  - Our approach for character segmentation needs lines as input, as a sequence of lines on top of each other breaks the algorithm.
- Character segmentation
  - We need to segment each character because the classifier cannot distinguish several characters from one image.
- Data formatting/casting
  - The data we want to test a classifier on needs to match the data we trained our classifier with. Hence we need to format our data to the same format as the datasets. Done using simple array manipulation, described in [2.3.4](#)

#### 2.3.1 Find rotation

##### Approach: OpenCV minAreaRect()

This approach uses `convex hull` to find the convex hull of the text, and then `rotating calipers` to find the minimum area rectangle.

`cv.minAreaRect` returns text rotated in one of following angles [0°, 90°, 180°, 270°], see Figure [3.2](#) for illustration. Hence we have a limitation on what angles the text can have, [0, 90] degrees, for us to be able to rotate and classify the text correctly. An approach that covers this vulnerability is mentioned in Chapter [3](#).

1. **Binary image.** For the Convex hull algorithm to work, the text segments and the background needs to be distinguishable. Convert image to a binary image using OpenCV threshold function and/or `bitwise_not` to flip foreground and background colors.
2. **cv.minAreaRect().** Feed our newly generated image to `cv.minAreaRect`.
3. **Calculate angle differences.** From `cv.minAreaRect()` we can find the angle of the rectangle. From that we can calculate the angle of the potential rotated rectangle.
4. **Affine transform to rotate image.** We use Affine transform to rotate the rectangle.

### 2.3.2 Find line

After all the preprocessing done up to this point, we assume we have a segment of correctly rotated text. At this point it is enough to just use projection histogram. We basically sum number of active pixels in each row and end up with one dimensional array with same size as image height, one value for each row in input image. E.g. [0 0 0 0 0 5 12 18 20 15 11 0 0 0 0 5 7 8..], this means there are some data (most likely text) between 5th and 10th row, and 15th up to another line break. see fig 2.1

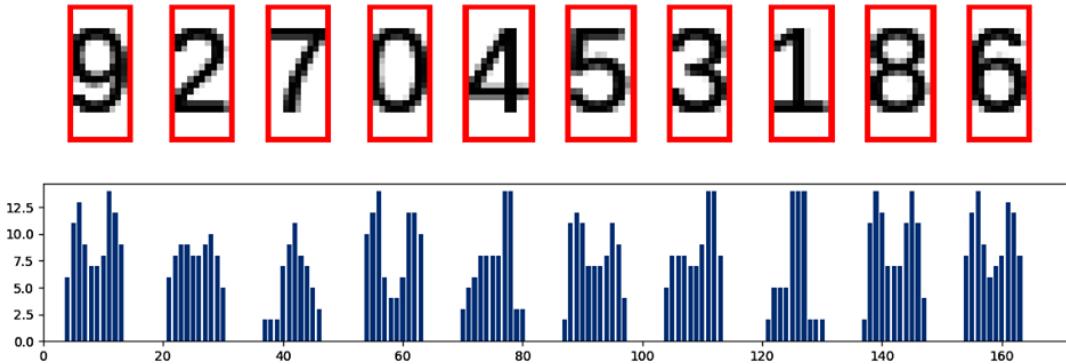


Figure 2.1: Vertical project histogram, same applies to horizontal

### 2.3.3 Find Symbol/Letter Segmentation

Letter segmentation was similar to Text Segmentation, section 2.2. We added a few steps and kernel shapes in the morphological part. The additional step was to fill holes in the image, for example; 8 and O, can give multiple wrong contours. cv2.floodFill was used to solve this problem.

1. Get binary image.
2. Do morphological dilation with a long vertical kernel, This is to include the dot in 'i'.
3. Add a border around the line-image, that separates elements from the edge after the previous step. This is to make cv2.floodFills work.
4. Fill all holes, cv2.floodFills.
5. Use cv2.FindContours
6. Ignore element where contour size smaller than a threshold(half the letter size), to ignore '.,.' and other non letter element.

### 2.3.4 Data formatting/casting

After we extract single symbols from the lines, we have to feed them to the classifier. As the classifier is trained on 28x28 grayscale images, we could simply resize the images, but this will more often than not lead to poor results as the images does not follow the same convention; the images used to train the network did, except for the image size ofc. So we have to do some manipulations in order to make our images look

more like the images our CNN is good at recognizing. Lets take a look at some images from our training sets:

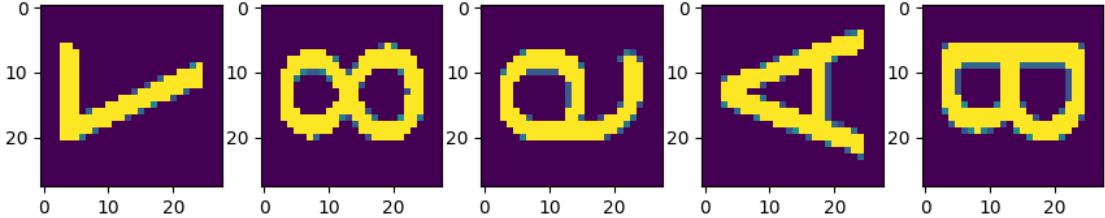


Figure 2.2: Dataset samples

As shownen above, the characters are centered and have about 4 pixels as border around them. Furthermore they are not binary images with sharp edges, but rather smoothed out. Additionaly the x and y axis are swapped, making the image look transposed, but that's how the EMNIST dataset has stored its data, so we just decided to stick with their convention.

The algorithm we found working best is described below:

1. We find the dominant axis of the image using  $n = \max(x, y)$ .
2. We create a new image with both axis equal  $n + n//k$  where  $k = 7$  as we found  $\frac{28_{total\_pixel}}{4_{border\_pixel}} = 7$ .
3. We paste our original character image in the center of our newly created background image, and now resize it to 28x28 pixels. The resize operation will smooth out the edges, even though the original image may have been a binary image.
4. Decided to remove unnecessary noise, the resize operation may have introduced by setting all pixels with intensity value under 50 to zero, and bring out the main shape of our character by setting all pixels with intensity value over 127 to 255 (this only affects the core of the character, edges are still smoothed). This should give us an image most likely to be correctly classified by our CNN. The final image is then transposed to match EMNIST data.

## 2.4 Component: Classification

### Description

For the classification component there were only 2 approaches we considered, Convolutional Neural Network (CNN), the architecture is illustration in Figure 2.3, and a Multilayer Perceptron (MLP or Deep neural network (DNN)), Figure 3.3 illustrates the architecture. However the method we ended up choosing for the final result is CNN. In Section 3.4 we explain why we discarded the MLP approach.

### Convolutional Neural Network

The basic idea of a CNN is; train some number of filters and then after the filters are trained, run the image which is now “filtered”, through a fully connected network. The fully connected network will then try and classify based on the feature images from the convolution layers.

Below we present the architecture and variable choices we have made for the CNN.

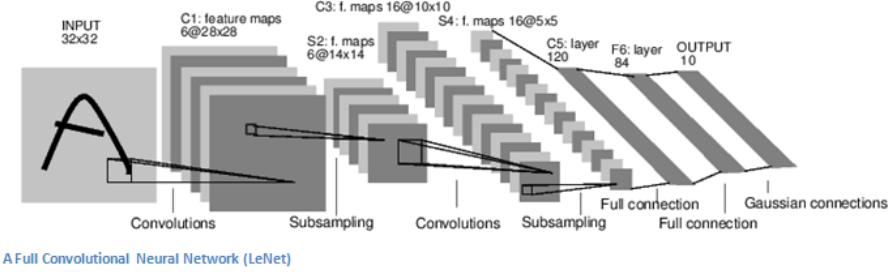


Figure 2.3: Convolutional Neural network [Source](#)

Our final model/topology looks like this:

$\text{conv}(32 \times 5 \times 5) \rightarrow \text{pool}(2 \times 2, 2) \rightarrow \text{conv}(64 \times 5 \times 5) \rightarrow \text{pool}(2 \times 2, 2) \rightarrow \text{conv}(64 \times 5 \times 5) \rightarrow \text{dense}(1024) \rightarrow \text{dropout}(0.4) \rightarrow \text{dense}(48)$

1. **conv(X×Y×Z)**: X-filters, Y×Z filter size, All conv layers have ReLu as an activation function.
2. **pool(2×2, 2)**: 2×2 pooling size, 2 stride. Standard size in smaller networks.
3. **dense(Z)**: Z outputs
4. **dropout(F)**: dropout rate (1-F probability that element will be kept)

Second and third conv-layers have 64 filters as we thought this amount is needed to recognize more complicated geometry in characters like **G**, **Q**, and maybe for further **Æ, Ø, Å**. We added pooling layers for two reasons: to reduce computation time and to add some rotation invariance. Dropout layer was added to reduce overfitting during training.

Given the time we had left for model testing this topology have us best result.

## 2.5 Component: Datasets

### 2.5.1 Description

In order to learn our network to distinguish between the characters it needs training. Training is done by feeding images of known objects to the network (labeled data) and telling it how inaccurate it's prediction so that network can adjust its weights accordingly. This type of training is called supervised training as we guide our network during training process. For network to get high accuracy of prediction a lot of labeled data is needed. Lucky for us datasets like MNIST exists, more info about it later.

#### description Limitation

As we have limited us to the digits and English alphabet, we will need labeled data for each of these [0..9] + 36 characters; divided into training, test and validation sets. As the concept of classifying only numbers vs all 36 characters does not differ that much, we will first see if we can solve the OCR problem with just numbers. Therefore we only need a dataset containing numbers at first. After that we can proceed our search for a dataset containing all the characters we need.

## Dataset

### SANS

Dataset created by ourself. Small dataset containing only 3 most used font in machine typing: **Sans**, **Times New Roman** and **Calibri**. These will be the foundation for our CNN as we for the most part will try to recognize characters from there fonts.

Therefore we decided to run multiple training steps on this dataset to learn our network the 'etalon' of each character.

### FNIST

Another dataset generated by our scripts. This one contain over 1000 free fonts found on the Internet. Here we can find examples of multiple variations of the same character like **bold**, *italics* and regular.

### EMNIST-ByMerge

The EMNIST dataset is a set of handwritten character digits derived from the NIST Special Database 19 and converted to a 28x28 pixel image format and dataset structure that directly matches the MNIST dataset. Further information on the dataset contents and conversion process can be found in the paper available at

<https://arxiv.org/abs/1702.05373v1>

# Chapter 3

## Discarded Method

### 3.1 Description

This chapter covers the methods tried but not included in the end result, because they showed dissatisfactory results. This chapter is also organized like Chapter 2, but for component description please refer to the previous chapter.

### 3.2 Text Segmentation

We tried to expand the original approach to be able to detect text with noise, natural images. We tried Stroke Width Transform first, then OpenCv scene text detection. We abandon both approach because of time constrain and difficulties of implementation. We decided therefore to focus on other parts of the project first.

#### Discarded Method 1: Stroke Width Transform

We tried Stroke Width Transform to do Text Segmentation when original approach gave a decent result. It was originally proposed by Epstein et al 2010 [3]. Since OpenCv does not have this implemented, we tried to implement it ourself. Additional sources was used in our attempt to implement it[1, 2, 6]. We were not able to finish this, but we think its worth mentioning, since we spent some time on this. The steps of Stroke Width Transform are as followed:

1. **Edge Detection and edge orientation(Done).** We need to have Edge image and orientation of the gradient image. Canny and Sobel was used in the original paper and other sources. This is simple since OpenCv has both Canny and Sobel implemented.
2. **Stroke Width Transform(Done).** Here we had to do more. We have to find a line from a starting point and the angle. We were able to implement this part, but there were some uncertainties. It only worked on black text with white background. That is because the orientation(Sobel filtering) is dependent on it. The paper talks about doing a second pass with inverse image, but we decided to ignore it in order to come farther in the algorithm.
3. **Find Connected Component(Done).** In this point we want to find connected components. The caveat here is that the components need to be connected with regards to the Stride Length. We used 'One Component At A Time' algorithm to find all the different components. This part we were able to finish.

4. **Exclude noise and find letters(not Done).** Since Stroke Width Transform tend to make a lot of noise. The obvious one is making single lines. This part is suppose to exclude this noise and at the same time exclude anything that is not a letter. The theory is, since letters and text in general, usually have the same stroke width, we can use this information do estimate what is a letter and what is not. We were not able to finish this part.
5. **Find lines/words(not Done).** Was not able to get to this part, but idealy it will combine letters to a single line or word.

In cases where the image has a lot of non text objects, it will work fine with it. We ended up discarding this approach since it was too time consuming and decided on working on a simple approach first.

### Discarded Method 2: OpenCv Scene Text Detection

OpenCV has its own Text Scene Detection. The approach of this algorithm is to detect text in scene using Classifier and Components Tree, propose by Lukás Neumann & Jiri Matas [5]. Since we already discarded Stroke Width Transform to focus on a simple approach, we decided not to use it. We had some problem to get proper result as well.

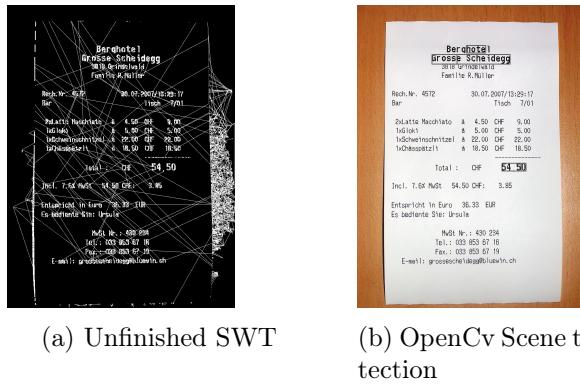


Figure 3.1: Result of Discarded Segment Text Method

## 3.3 Preprocessing

We tried different approaches only on find rotation and character segmentation. Line segmentation and dataset matching we didnt try other methods on.

### 3.3.1 Find rotation

The original thought was to use the Hough Transform to find the lines in a text, and then from that, find the rotation. However because of bad results, we moved to cv.minAreaRect. This too had some limitations on what it could rotate, see Section 2.3.1. These limitation we tried to solve using a CNN solution, presented below.

### Discarded Method 1: Hough Transform

**Hough transform** is a well known algorithm to find lines, its approach is to see if it can aligne some threshold of pixels on one straight line. To help with this process we will

do a simple edge detection algorithm on the image before running it through the Hough Transform.

Below, the steps needed to perform this approach are mentioned.

1. **Edge detection** To make the Hough transform perform better we want to remove unnecessary noise. Canny edge detection algorithm is a robust and fast solution for this.
2. **Line detection** Now perform the actually Hough Transform.
3. **Rotate image** Lastly we need to rotate the image according to the lines from the Hough Transform.

#### **Discarded Method 2: OpenCV minAreaRect() + CNN solution**

This approach builds on the approach presented in Section 2.3.1. As cv.minAreaRect returns text rotated in one of following angles [0°, 90°, 180°, 270°], see Figure 3.2 for illustration. We wanted to add a CNN to determine which of the angles is correct and then rotate it accordingly. To train the network we needed a dataset with the corresponding angles as labels, as we could not find one, we decided to create it ourselves, see Section 3.5 for more on this.

We discarded this approach because it gave only a 65% accuracy of the right angle, and we are only working on simple text images with little rotation, hence no need for this addition in our software.

The approach is described below:

1. **OpenCV minAreaRect()** The same approach mention in section 2.3 find rotation
2. **CNN solution - find correct rotation** Now that we have text rotated in one of [0°, 90°, 180°, 270°], we have several options finding the correct angle of the text segment.
  - (a) The most straightforward approach would be to find the rotation of first character and rotate the whole segment accordingly (fast and naive)
  - (b) Pick N characters and find their rotation, angle with most matches will be our final angle of rotation. This is somewhat slower (depends on N) but gives us bit more confidence than (a)
  - (c) Finally we can classify the angle of each individual character before we try to determine the correct angle based on the results. This gives us the highest accuracy but on the cost of time/computation power. This is the only approach able to successfully recognize this type of text: (find better example) **LATEX**



Figure 3.2: `cv.minAreaRect` cannot differentiate between  $0^\circ$  and  $180^\circ$ , and  $90^\circ$  and  $270^\circ$

### 3.3.2 Character Segmentation

Since we used Projection Histogram to find lines, we can use it to find character segments as well. We started initially to do that, but later switched to the new approach.

This approach is similar to Line detection in Section 2.3.2. Its weakness is that we may not get the right height of each character, See figure 2.1, meaning we get more background. This is bad because later to match the desired format for classification, the resizing/reshaping makes these empty spaces even larger, and the letters even smaller. Consequently it gets more difficult to classify correctly.

## 3.4 Classification

As mentioned in Section 2.4, we also tried to use MLP to do classification of characters. We decided to avoid using MLP in our project as it required lots of training data; it is not able to detect spatial or minor rotation changes, e.g. a network trained with MNIST with accuracy over 95% was doing poorly in recognizing machine-printed characters, even though they were etalon of a perfect input.

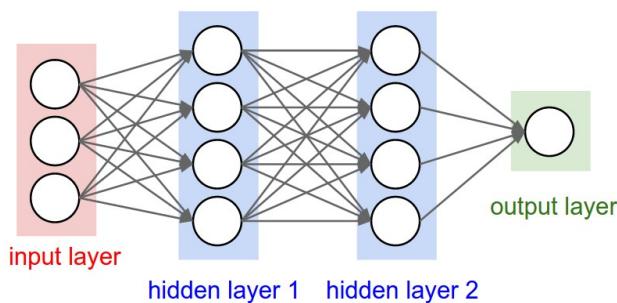


Figure 3.3: MLP Neural network [Source](#)

## 3.5 Datasets

### 3.5.1 MNIST

*MNIST* was discarded as we decided to stick with machine printed digits and letters, none of which were included in MNIST. It was a good starting point to test our networks (both MLP and CNN), however to be able to achieve a classifier which could classify the English alphabet in addition to numbers ranging from [0-9] we had to go with EMNIST.

### 3.5.2 ROT\*

*ROT\** - dataset containing rotated examples of the same dataset, but with labels corresponding to angle of rotation, e.g. [0, 90, 180, 270].

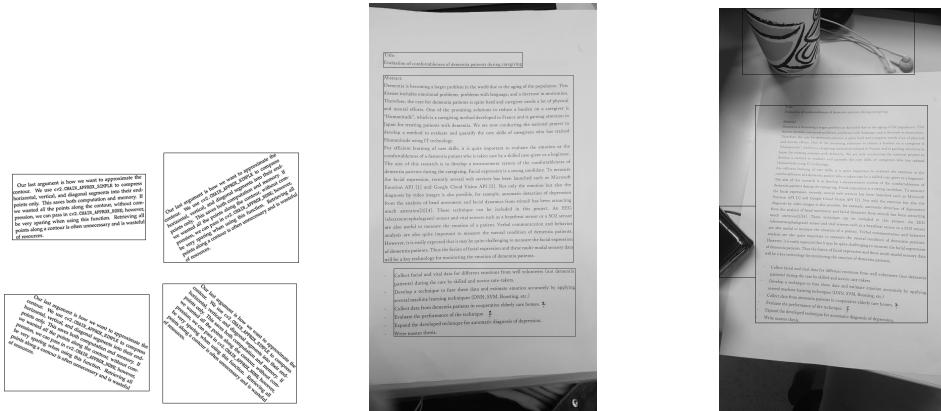
This dataset was discarded as our network showed very little learning progress as the accuracy of predictions was constantly around 65%. This is not the datasets fault, most likely it was due to bad network topology or poor choice of hyper parameters but we didn't have enough time to debug this part of the project and decided to stick to OpenCV's minAreaRect() results.

# Chapter 4

## Result - Discussion

### 4.1 Result: Text Segmentation

We tried 3 Approaches, Morphological, Stroke Width Transform and OpenCv Scene text detection. We ended up with the simple solution, Morphological approach. It gives good results, but has a lot of limitations. It only works on black text with white background, images with little noise, and with text only in the image, you can see the result in figure 4.1. We could include a part to distinguish between text and non-text areas to improve this solution. A strategy to do that can be to compare numbers of connected components, since we know text regions have a lot of letters/components, compared to a coffee cup or a wallet.



(a) Good result on pure text  
(b) Good result on text on paper with little background noise  
(c) Bad result with many non-text objects segmented as text

Figure 4.1: Result of the Text Segmentation

### 4.2 Result: Preprocessing

#### Rotation

Rotation only work if the angle of rotation is less than 45 degree. This problem is mentioned in section 2.3, and an attempt to solve the problem was given in section 3.3.1. Below the result of a image with rotation less than 45 degrees is showed in figure 4.3

Our last argument is how we want to approximate the contour. We use cv2.CHAIN\_APPROX\_SIMPLE to compress horizontal, vertical, and diagonal segments into their endpoints only. This saves both computation and memory. If we wanted all the points along the contour, without compression, we can pass in cv2.CHAIN\_APPROX\_NONE; however, be very sparing when using this function. Retrieving all points along a contour is often unnecessary and is wasteful of resources.

(a) Region of skew text, after text segmentation

Our last argument is how we want to approximate the contour. We use cv2.CHAIN\_APPROX\_SIMPLE to compress horizontal, vertical, and diagonal segments into their endpoints only. This saves both computation and memory. If we wanted all the points along the contour, without compression, we can pass in cv2.CHAIN\_APPROX\_NONE; however, be very sparing when using this function. Retrieving all points along a contour is often unnecessary and is wasteful of resources.

(b) After rotation

Figure 4.2: Result of rotation

#### 4.2.1 Line segmentation

If the image has no space between lines or is slightly rotated, it can cause an error, lines will not be separated. As our previous steps handles these problem our method worked quite well on our test images.

**Lorem Ipsum** is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

(a) Text Region after rotation

**Lorem Ipsum** is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

(b) Boundary box of line region

Figure 4.3: Result of line segmentation

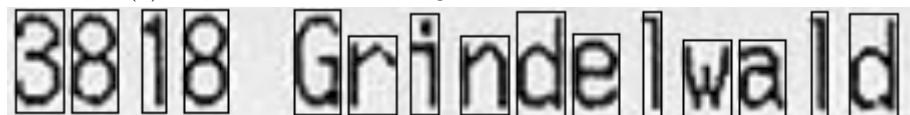
#### 4.2.2 Character segmentation

This part worked quiet well, an alternative would be to just use OpenCv findComponent. The steps and results would not differ too much, but we could have dropped cv2.floodFills.

A weakness of this approach is, if multiple characters are merged with one another we would not be able to separate them. A solution for this problem can be to do a morphological opening to separate character better. Another problem we came across was we require image to be in a somewhat good quality. If image have not good enough quality it became harder to separate character from each other, as we require clear space in between them.

**Lorem Ipsum is simply dummy text of the printing and**

(a) Result of Character Segmentation from extracted line



(b) Result of Character Segmentation from an example line image

Figure 4.4: Result of Character segmentation

## 4.3 Result: Classification

### 4.3.1 Description

Convolutional neural networks are especially good for image classification, because they take local spatial connections into account when they classify. This way it doesn't matter where in the image our object/character is it will be able to recognize it, same yields for rotation, as the CNN classifies based on local spatial connections it doesn't matter if the object is slightly rotated. Hence the classification would be even more robust compared to the MLP.

Following figures show topology and performance during training. Topology stays the same as described in Figure 4.5. In Figure 4.6a we can observe how the accuracy changes over time. You can notice drastic acceleration after 2'000 steps, this is due to us changing the dataset context as we used multiple independent datasets during one training session. Loss graph 4.6b swings follow expected path and resemble pretty much every loss graph for correctly setup network, be it MLP, CNN or any other. The final accuracy we were able to get is **87.76%** and it only took us approximately 30 min of training on **nVidia GeForce 840M GPU**.

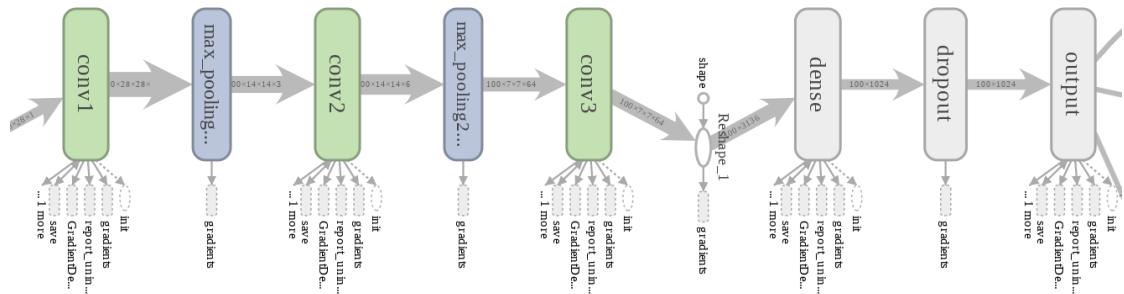


Figure 4.5: CNN topology

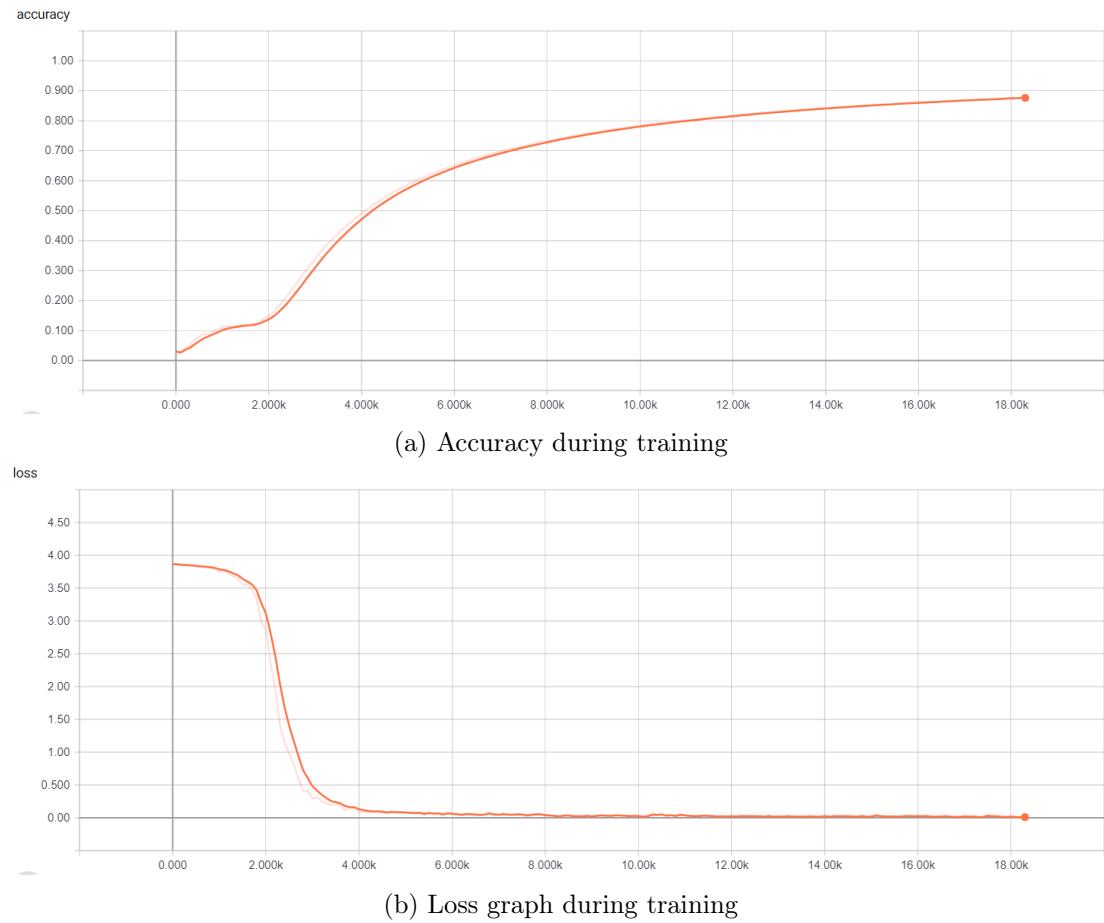


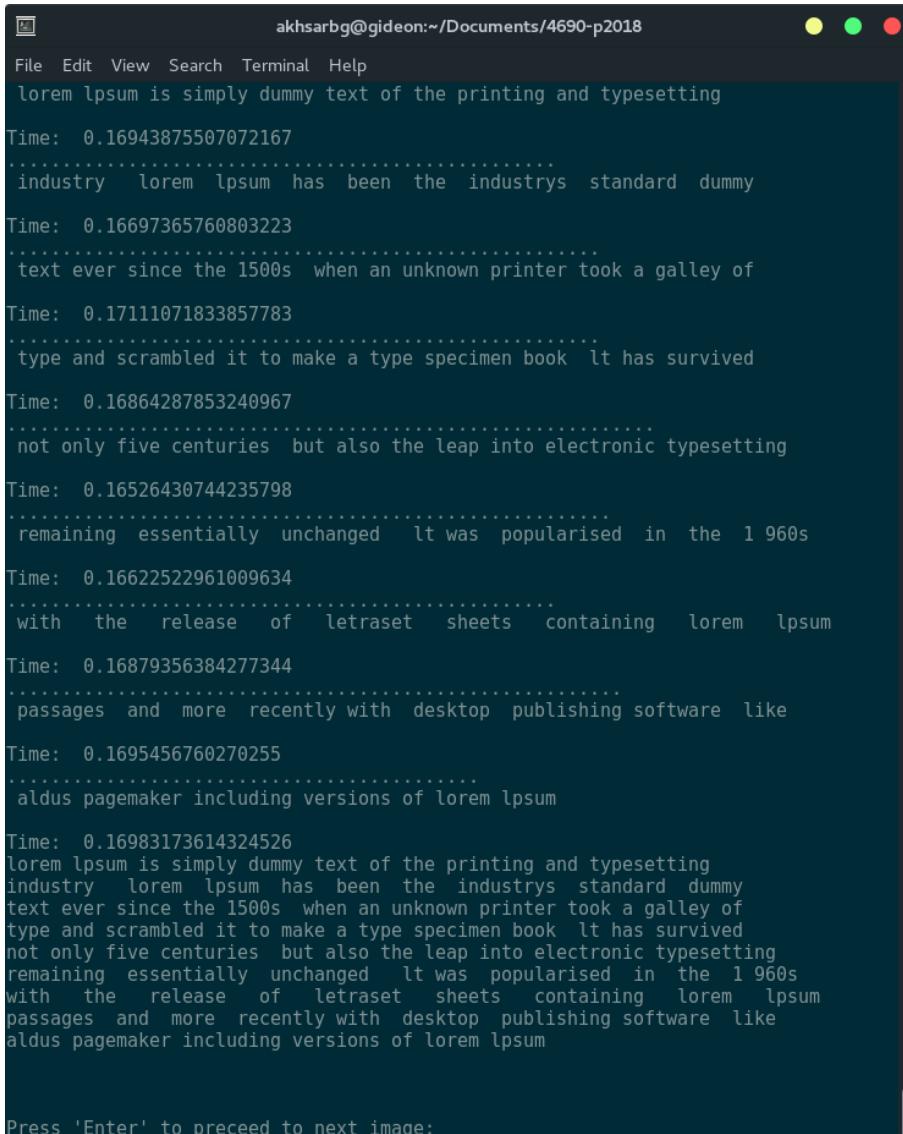
Figure 4.6: Result of Classification

## 4.4 Final Result

We saw that each component was able to its part reasonably. Each have weakness and fails, but if we are to combine them together we are able to do get a somewhat reasonable result. Figure 4.7 show an input image and what we got in the output. Some of the letter are not able to be classify correctly. The main error came from character segmentation and classification, but we got somewhat readable result...

**Lorem Ipsum** is simply dummy text of the printing and typesetting industry. Lorem Ipsum has been the industry's standard dummy text ever since the 1500s, when an unknown printer took a galley of type and scrambled it to make a type specimen book. It has survived not only five centuries, but also the leap into electronic typesetting, remaining essentially unchanged. It was popularised in the 1960s with the release of Letraset sheets containing Lorem Ipsum passages, and more recently with desktop publishing software like Aldus PageMaker including versions of Lorem Ipsum.

(a) Original image 'lorem.png' as input image



```
akhsarbg@gideon:~/Documents/4690-p2018
File Edit View Search Terminal Help
lorem ipsum is simply dummy text of the printing and typesetting
Time: 0.16943875507072167
industry lorem ipsum has been the industrys standard dummy
Time: 0.16697365760803223
text ever since the 1500s when an unknown printer took a galley of
Time: 0.17111071833857783
type and scrambled it to make a type specimen book lt has survived
Time: 0.16864287853240967
not only five centuries but also the leap into electronic typesetting
Time: 0.16526430744235798
remaining essentially unchanged lt was popularised in the 1 960s
Time: 0.16622522961009634
with the release of letraset sheets containing lorem lpsum
Time: 0.16879356384277344
passages and more recently with desktop publishing software like
Time: 0.1695456760270255
aldus pagemaker including versions of lorem lpsum
Time: 0.16983173614324526
lorem lpsum is simply dummy text of the printing and typesetting
industry lorem lpsum has been the industrys standard dummy
text ever since the 1500s when an unknown printer took a galley of
type and scrambled it to make a type specimen book lt has survived
not only five centuries but also the leap into electronic typesetting
remaining essentially unchanged lt was popularised in the 1 960s
with the release of letraset sheets containing lorem lpsum
passages and more recently with desktop publishing software like
aldus pagemaker including versions of lorem lpsum

Press 'Enter' to preceed to next image:
```

(b) Terminal output of lorem.png

Figure 4.7: Final Result of lorem.png

# Chapter 5

## Conclusion

### 5.1 Time pressure

Unfortunatly we encounterd a relativly large time pressure problem on the report. There are several stuff we would like to have had the time to comment on, such as; add our thoughts on future improvments, comment all the weaknesses we discovered in the final parts of the software (which we regret not being able to comment on), and ofcourse, proof read the entire document.

### 5.2 Ambition vs reality

Our goals was relatively ambitious compared to the end product. We envisioned a software which was rotation invariant, light intensity invariant to some extend; could recognize text with great accuracy. However, it was a far more challenging task then what we expected.

Firstly we started off with an MLP network, which is not the best choice for image classification. Secondly we didn't have much experience with machine learning software, so most of our solutions were combinations of trying to see what worked for others and researching what could give a good result for our problems. Additionally, as writing everything in the python environment would be very slow, we decided to write the machine learning algorithms with the TensorFlow API. This too was a challenge, as it takes some time to understand how everything in the TensorFlow environment works.

### 5.3 Final remarks

Challenge wise, a really great problem, learned a lot about TenserFlow, OpenCV, CNN, MLP, different datasets available open source.

Thought the time pressure in the end got us really worked up, without it we doubt we would of worked as consistently with the project as we did!

Considering everything mentioned above, this project was one of the best “assignments” we have done at IFI/UNIK.

# Chapter 6

## References

### 6.1 Note

Unfortuntly because of time pressure, and other exams we didnt have the time to make this part more presentable.

Here is the list of online sources we used for the different parts of the project. We probably used more, but these are what we recognize as heavily influenced online sources:

#### 6.1.1 Text Segmentation

- <https://www.danvk.org/2015/01/07/finding-blocks-of-text-in-an-image-using-python-opencv.html>
- <https://github.com/mypetyak/StrokeWidthTransform>
- <https://stackoverflow.com/questions/24385714/detect-text-region-in-image-using-opencv>
- [https://github.com/opencv/opencv\\_contrib/tree/master/modules/text/samples](https://github.com/opencv/opencv_contrib/tree/master/modules/text/samples)

#### 6.1.2 Preprocessing

- <https://www.pyimagesearch.com/2017/02/20/text-skew-correction-opencv-python/>

#### 6.1.3 Classification

- A Guide to TF Layers: Building a Convolutional Neural Network
- The EMNIST Dataset

#### 6.1.4 OpenCv tutorials

We used OpenCv a lot and found many of the tutorials helpful. Here is a list of few of them, but we may have used more.

- [https://docs.opencv.org/3.0-beta/doc/py\\_tutorials/py\\_imgproc/py\\_contours/py\\_contour\\_features/py\\_contour\\_features.html#contour-features](https://docs.opencv.org/3.0-beta/doc/py_tutorials/py_imgproc/py_contours/py_contour_features/py_contour_features.html#contour-features)
- [https://docs.opencv.org/trunk/d9/d61/tutorial\\_py\\_morphological\\_ops.html](https://docs.opencv.org/trunk/d9/d61/tutorial_py_morphological_ops.html)

# Bibliography

- [1] Christopher Bunn. *StrokeWidthTransform: A playground implementing Epshtein, Ofek, and Wexler's Stroke Width Transform.* original-date: 2015-11-18T04:42:20Z. May 8, 2018. URL: <https://github.com/mypetyak/StrokeWidthTransform> (visited on 05/24/2018).
- [2] c++ - Given an angle and length, how do I calculate the coordinates. Stack Overflow. URL: <https://stackoverflow.com/questions/1638437/given-an-angle-and-length-how-do-i-calculate-the-coordinates> (visited on 05/04/2018).
- [3] Boris Epshtein, Eyal Ofek, and Yonathan Wexler. "Stroke Width Transform." In: IEEE, Jan. 2010. URL: <https://www.microsoft.com/en-us/research/publication/stroke-width-transform/>.
- [4] Finding blocks of text in an image using Python, OpenCV and numpy. URL: <http://www.danvk.org/2015/01/07/finding-blocks-of-text-in-an-image-using-python-opencv-and-numpy.html> (visited on 04/26/2018).
- [5] L. Neumann and J. Matas. "Real-time scene text localization and recognition." In: IEEE, June 2012, pp. 3538–3545. ISBN: 978-1-4673-1228-8 978-1-4673-1226-4 978-1-4673-1227-1. DOI: [10.1109/CVPR.2012.6248097](https://doi.org/10.1109/CVPR.2012.6248097). URL: <http://ieeexplore.ieee.org/document/6248097/> (visited on 05/21/2018).
- [6] Gili Werner. *Text Detection in Natural Scenes with Stroke Width Transform.*