1. OCR Intro
2. OCR History
3. OCR Application
4. Why OCR is important
5. Creating OCR
6. MNIST Dataset
7. OCR Architecture
8. GAN (Generative adversarial network)
9. Hieroglyphics
10. Character Recognition

## Optical Character Recognition

#### Intro

What is OCR? OCR or Optical Character Recognition is the process that converts an image with text into a machine-readable text format, such as a word document, or a txt file, but how do they work exactly? Before we answer this question let’s take a look of the history of OCR.

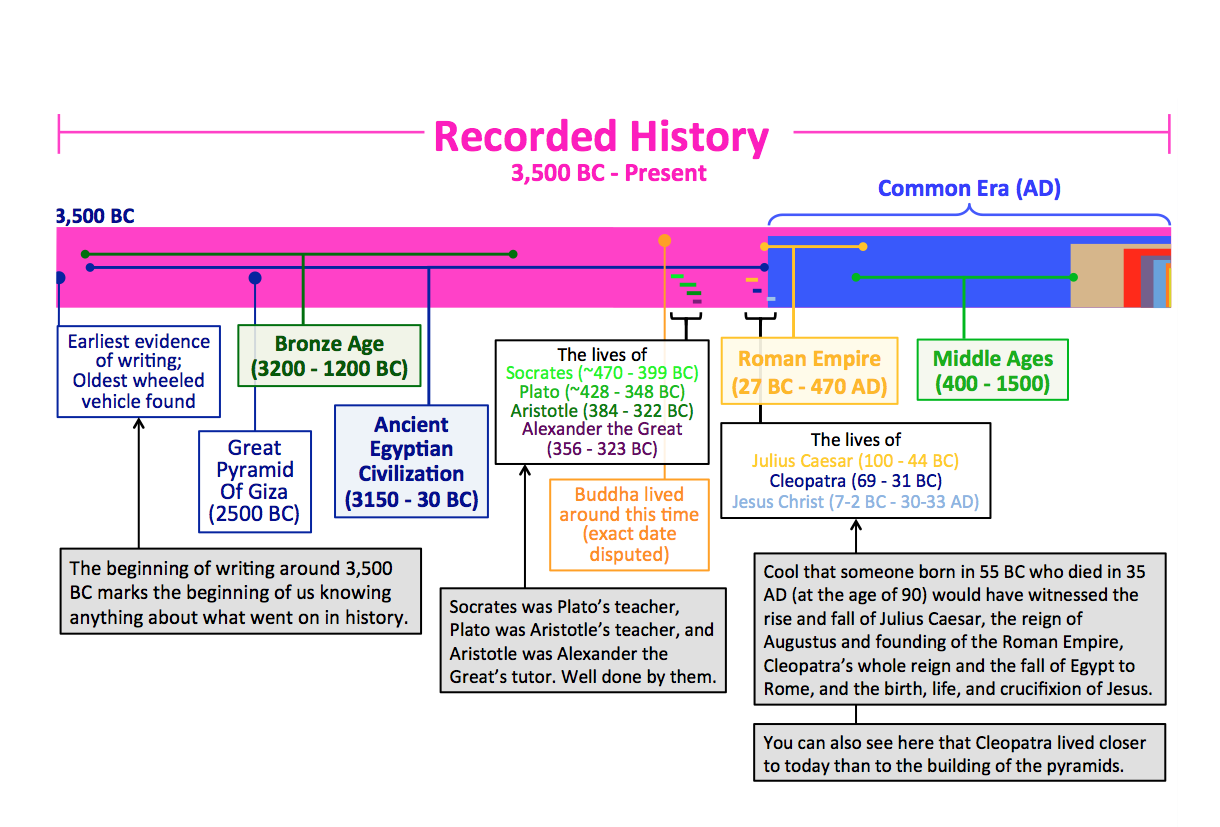
### History

Many people may think that OCR has only been around for a decade or so, but it turns out that they have been here for way longer than you think.

This is an image of an Optophone from 1914, which can read characters by using selenium photosensors to detect black print and converting them into audible output. This technology has since been improving for hundreds of years, and has developed to the point where this analog-to-digital documentation method reached one of its milestones in recent years.

### Importance

The question arises, “Why OCR is important for humanity?” One of the major reasons, is to encourage data preservation.



Take a look at the recorded history of mankind. It has spanned over 5500 years until only the recent 2-3 decades are when data is preserved in the forms of 1s and 0s. As such, it is impossible to hand-document human history by copying data from stone walls to the keyboard, one by one. This is where OCR technology comes into play.

### Application

With the help of OCR, not only can ordinary people use it to save data in a faster manner, but also can historians, archaeologists, and researchers benefit from this. Storing data in a binary form can allow us to search for data and citations efficiently, which signifies the importance of the ability to convert handwritten characters into digital copies.

## Creating OCR

#### MNIST Dataset

During this course, I spent the majority of my time researching Chinese OCR models. And before we start tackling the big boss, let’s try making a model which can recognize handwritten digits. In this example, I’ve used TensorFlow to achieve this. For those who want to follow my steps, I’ve used their MNIST dataset, which includes a whole library of handwritten digits, to train a model which can identify these images. After about 20 seconds of training, the results are here. As you can see from this graph, the model is slowly learning the patterns of the dataset, and can already have a 97% accuracy with this bear amount of preparation. And as such, we have created the most basic framework of an OCR, which is the baseline of every character recognition module you see elsewhere.

#### OCR Architecture

However, as we delve deeper and deeper into the rabbit hole, we would find that the neural network that we currently are using is not enough to identify character sets that are so large that not even the locals recognize all of them. Take a look at languages like Chinese, which includes over 100 thousand different characters, which basically takes forever to train in a regular convolutional neural network.

To create a suitable Chinese OCR model, choosing the right architecture for our network is crucial to leverage the features of the samples .

Currently, mainstream OCR models use a two-part process, which are the text detection module, and a text recognition module.

<https://www.infrrd.ai/research/blog/transformer-based-ocr>

At the start, the program throws the image into a text detection module, such as YOLO, Detectron, or Mask-RCNN, in order to detect the bounding boxes for all the texts in the image. Afterward, a text recognition module is used to understand the content of the detected text block and convert the visual signals into natural language tokens.

In recent years, there is even a new type of neural network architecture which is known as Vision Transformers, and is known for being convolution free and does not rely on any complex pre or post-processing steps. However, transformers require a very large amount of training data and require a very large budget to achieve.

#### GAN

Another topic that has been in the spike of OCR, is Generative Adversarial Networks, also known as a GAN.

Assume that we are making a model for hieroglyphics. The problem arises when we are trying to collect enough data for training the network, as there is not enough sample data out there. This is where GAN comes in. Unlike a traditional neural network, which compares its outputs with the expected outputs, a generative adversarial network has two sub-networks, which are the generator and the discriminator. The generator would try to generate fake samples to feed it to the discriminator, and the discriminator would give an output of whether it is a real sample, or a fake sample. This way, the generator would try to generate samples which look real enough, that it can fool the discriminator, and eventually fool us humans as well.

Applications of GAN are seen in examples like Deep fake technology, and audio synthesis. With this technology, we can create larger datasets by providing a small data input, which further boosts the accuracy of the network. Of course, it still requires a human to identify whether the model is providing correct data in the first place, or else it would just be garbage-in garbage-out.

<https://en.wikipedia.org/wiki/Generative_adversarial_network>

<https://www.youtube.com/watch?v=TpMIssRdhco&ab_channel=IBMTechnology>

#### Chinese Character Recognizing Approach

For Chinese Character Recognition, the problem of training such a model is that the Chinese language contains a large amount of characters, and it would require too much time and effort just to put together a suitable dataset and train it. This is why I have 2 main ideas on how I can train such a model. Plan A, is to train a model to recognize Changjie Characteristics, but I soon realized I couldn’t find a suitable Python library to convert Changjie into Chinese Characters, so that method is put to the shelf for now. Plan B, is what I am planning to go for, which is to lay out a model which can recognize the eight strokes of Chinese Characters, which will be the second stage of my project. Hopefully, in the future four years of university, I can acquire more knowledge in this field of study, and complete my research on this topic. Thank you!