

HANDWRITTEN CHARACTER RECOGNITION

GROUP MEMBERS:

S.No	Name
1	Kurapati Sarayu
2	Surya Narayana Pasumarthi
3	Gangam Rushik Reddy
4	Naga Raviteja

INTRODUCTION & MOTIVATION:

Character recognition is a process which allows computers to recognize written or printed characters such as numbers or letters and to change them into a form that the computer can use. Handling large amount of manual files are not an easy task, the room full of document that are very important and may be necessary to be looked into any time. To search a file from the room full of files sounds like a hectic job but this is the situation on most of the government offices in our country and many other sectors and companies. A person looking for a file may lose temper may not be able to find the file at all or may spoil other file during the search due to frustration, all these are possible since it's a human searching for a file in the large pile of files. This intuition helps us realize the necessity of a character recognition library in every office. The library can be used to convert all the files and documents into electronic form which can be stored in a hard drive which consumes way less amount of space to store all the files. The file stores in the disk are easy to search since there exists so many tools to search through the files hence we have a whole room empty which can be used for various different works also reducing the stress of the person who is trying to search for a file. The

files which could be damaged or destroyed by mites will be protected from any physical damage. The files will also be portable easily and can have many copies of the files if it is needed to be present in different places at the same time. The data obtained from the files can also be used for data analysis and data interpretation for other research and study.

In these times, there is a great requirement of this kind of a recognition system. In today's world, where our main focus or emphasis is on employing efficient and cost-effective methods, many people are now using technology to perform tasks that have previously been laborious as well as time consuming. The essence of our modern gadgets is their efficiency as well as their role in reducing labour intensity. Computers are now used as a well-ordered and well regulated alternative for taking and managing notes and for eradicating potential issues regarding clarity of handwriting or misplacing sheets of papers.

Additional advantages of preferring technology involve the comfort of making such notes better at a later time, as well as reducing the wastage of physical space. One more example regarding the benefits of the present technology is utilization of tablets for both professionals and students, conserving the need for several books and various important documents, once more preventing wastage of extra physical space and letting all necessary data be kept inside a small piece of equipment. Data management has gained considerable prominence in engineering executions. Given that the industry sector is consistently increasing throughout the past several decades, a significantly greater demand over automated machinery can be observed. These automated machineries depend heavily towards data management in the form of character identification utilities. A simple example for such can be comparing data management akin to organization of mails at a relevant distribution complex, under which all data is processed by OCR platforms. Then, decisions are undertaken through automated machinery arms to sort the correct mail towards the correct lane so that all mails are sorted in relevance to the areas they need be delivered.

WORK DONE:

- Created a model to recognize the handwritten text.
- Developed a model to predict the accuracy

APPLICATION:

ANN Project examination - X training.ipynb - Colaboratory X +

colab.research.google.com/drive/1tqZe4lRtGX3vqGrZkNFygg1j4_0s-38

Flipkart Amazon.in AliExpress

training.ipynb ☆

File Edit View Insert Runtime Tools Help Last saved at 2:50 AM

+ Code + Text

Connect Editing

```
import numpy as np
import matplotlib.pyplot as plt
from extract import Extract
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPool2D, Dense, Dropout, Flatten
import tensorflow.keras as ks
%matplotlib inline
```

[] #Extracting data
(x_train,y_train),(x_test,y_test) = Extract('A_Z_Data.csv',0.2)

x_train = x_train.reshape(len(x_train),28,28,1)
x_test = x_test.reshape(len(x_test),28,28,1)

Dividing data into 297961 training and 74491 testing samples..

[] labels=['A','B','C','D','E','F','G','H','I','J','K','L','M','N','O','P','Q','R','S','T','U','V','W','X','Y','Z']

[] #lets display some data
for i in range(8):
 plt.subplot(2,4,i+1)
 plt.imshow(x_test[i].reshape(28,28),cmap='binary')
 plt.title(labels[np.argmax(y_test[i])])

T N O U

ANN Project examination - Xtraining.ipynb - Collaborat...

colab.research.google.com/drive/1tqZe4IRtGX3vqGrZkINfyyg1j4_0s-38#scrollTo=EciNzmeal58o

FlipkartAmazon.inAliExpress

training.ipynb

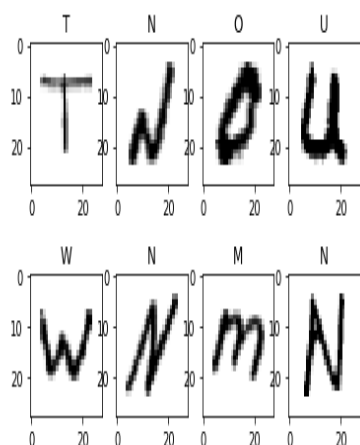
CommentShareSettingsS

File Edit View Insert Runtime Tools Help

+ Code + Text

Connect Editing

```
# display some data
for i in range(8):
    plt.subplot(2,4,i+1)
    plt.imshow(x_test[i].reshape(28,28),cmap='binary')
    plt.title(labels[np.argmax(y_test[i])])
```



```
[ ] #designing the model
model = Sequential()
model.add(Conv2D(filters=16, kernel_size=5, strides=1, padding='same', activation='relu', input_shape=(28, 28, 1), name='conv1'))
model.add(MaxPool2D(pool_size=(2, 2),name='pool1'))
model.add(Dropout(0.2,name='dropout1'))
model.add(Conv2D(filters=32, kernel_size=2, strides=1, padding='same', activation='relu',name='conv2'))
model.add(MaxPool2D(pool_size=(2, 2),name='pool2'))
model.add(Flatten(name='flat'))
```

ANN Project examination - Xtraining.ipynb - Colaborat... X

colab.research.google.com/drive/1tqZe4IRtGX3vqGrZkINFyyg1j4_0s-38#scrollTo=EciNzmeal58o

FlipkartAmazon.inAliExpress

training.ipynb

CommentShareSettingsS

File Edit View Insert Runtime Tools Help All changes saved

+ Code + TextConnectEditing

```
[ ] plt.plot(history.history['loss'])
plt.plot(history.history['val_loss'])
plt.title('Model loss')
plt.ylabel('Loss')
plt.xlabel('Epoch')
plt.legend(['Train', 'Test'], loc='upper left')
```

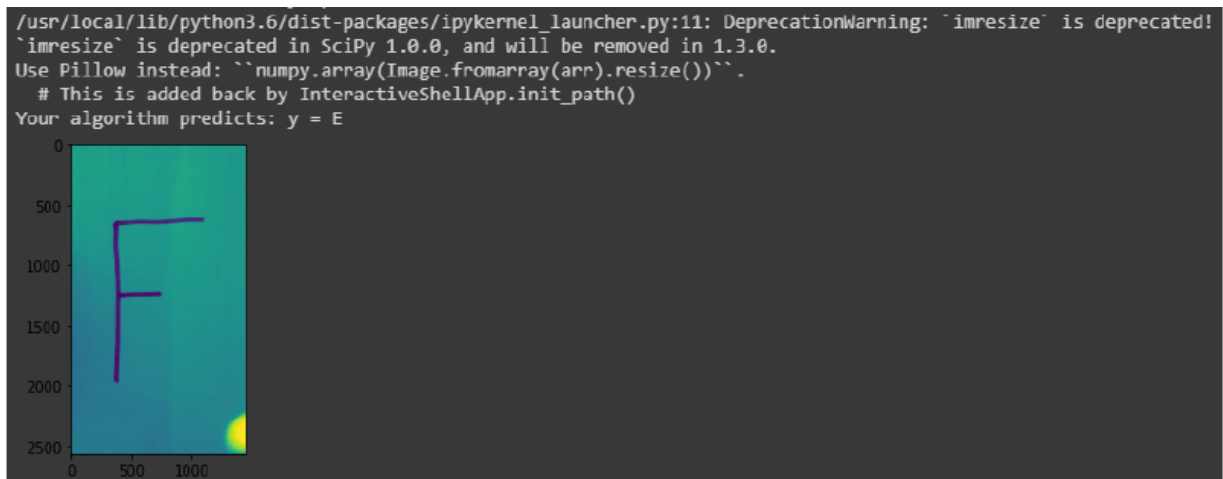
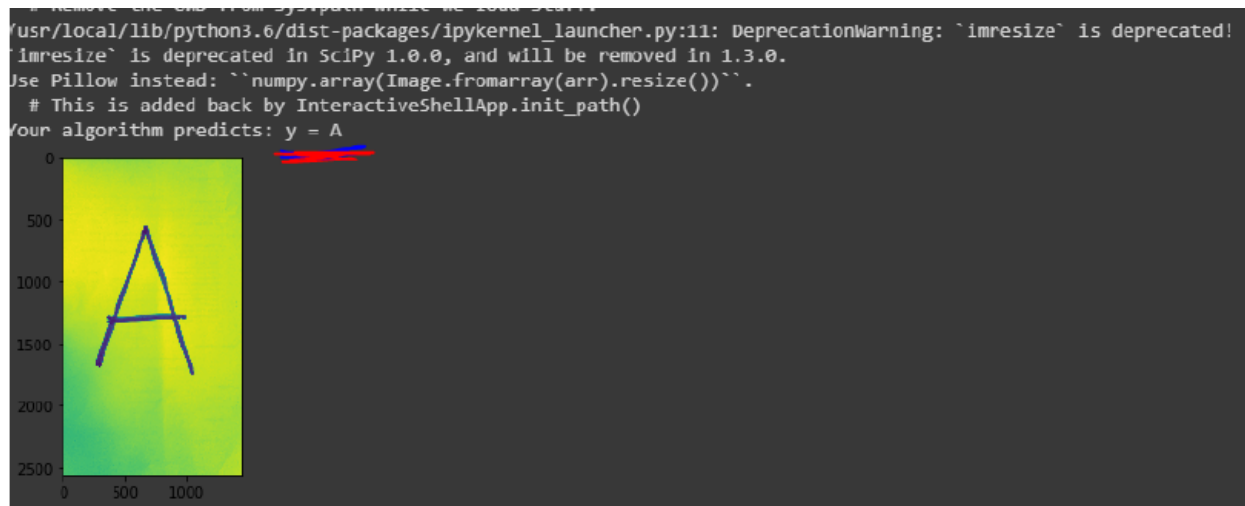
<matplotlib.legend.Legend at 0x179c9c54548>

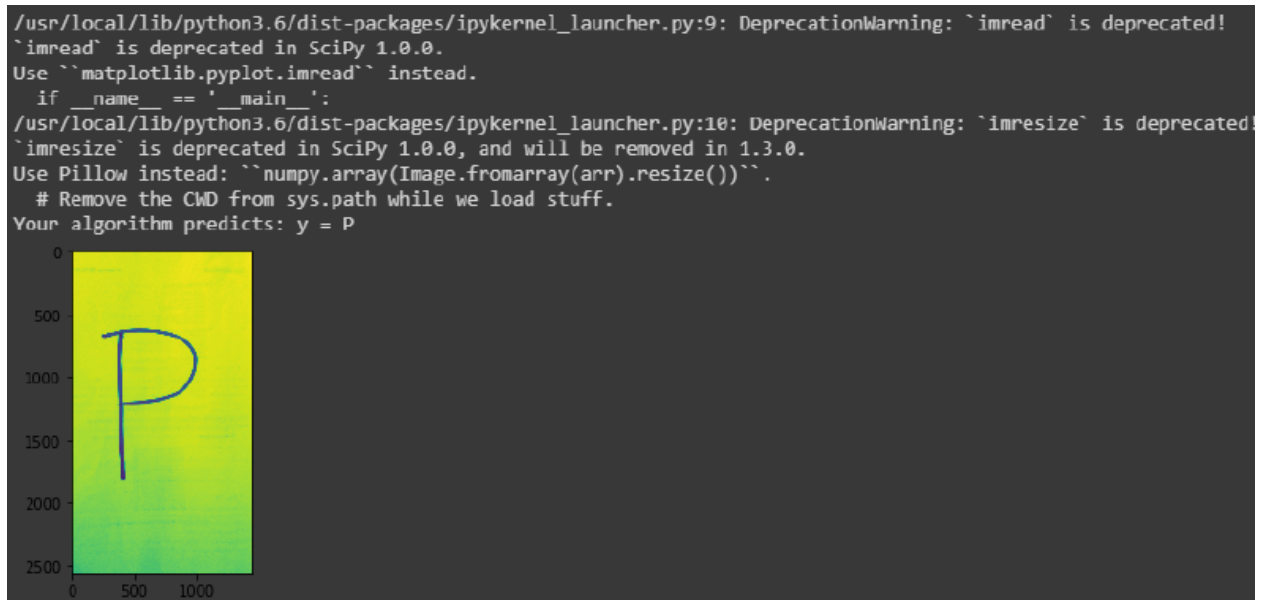
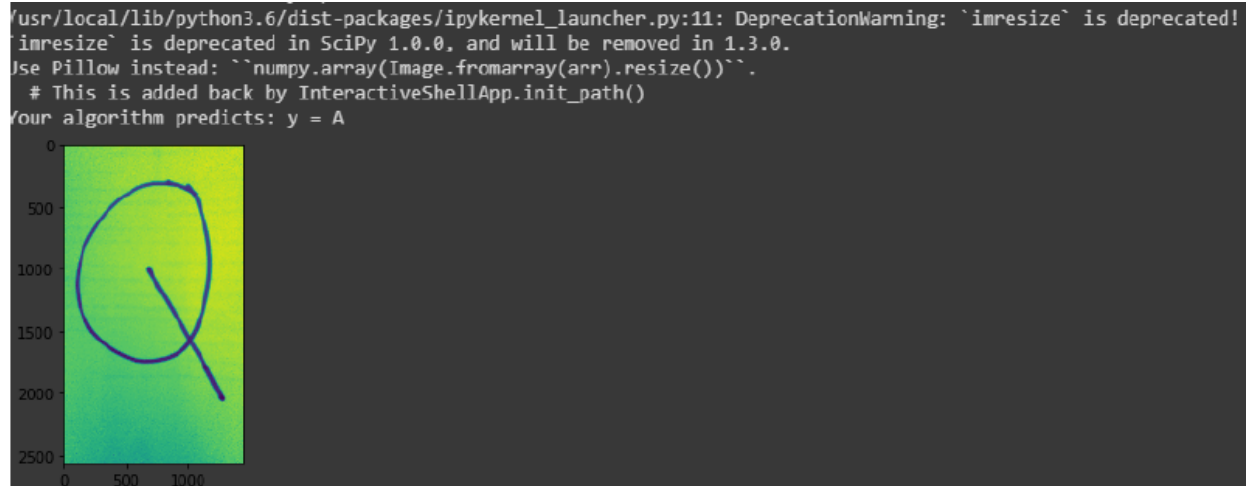
Epoch	Train Loss	Test Loss
0.0	0.55	0.18
2.5	0.12	0.08
5.0	0.08	0.06
7.5	0.06	0.05
10.0	0.05	0.04
12.5	0.04	0.04
15.0	0.04	0.04
17.5	0.04	0.04

```
[ ] model.save("model.h5")
```

```
[ ]
```

RESULTS :





CONCLUSION:

The structure behind identification of physical writing characters and building a python library function has been exhibited through this project. The span of this project involved functions i.e. acquisition of images, selection of characters, categorization, pre-processing process, and creation of outcomes. A study was furthermore conducted with the effect to evaluate the structural performance. Various developmental aspects are worth providing citation. The effectiveness of an OCR platform is based on the rate of conversion and accuracy. The accuracy is based on the volume of correct predictions establishing amidst the testing

procedure. Nevertheless, the rate of conversion is speed, over which the extraction of data was effectively executed. Under this project, a procedure has been recommended, alongside a model as implemented to effectively take out information and exhibit it. Several hindrances made the entire procedure more difficult, pertaining mostly towards misperception amidst images, i.e. D and O, identical features amidst letters, i.e. l or I, where l is L and I is i. Another major issue was the limited volume of test images. The tested material constituted our personal hand writing. Any differentiation over the style of physical writing could lead to poor accuracy in the recognition of input. All individuals possess unique handwriting and the software simply facilitates the infinite array of various writing styles. Printing quality, too, has a major influence over the extraction of data, i.e. data under black ink, featuring a black background will be difficult to identify, likewise, information which was exhibited over poor writing, i.e. writing from a young child, or someone lacking basic writing skills, would fail to provide the desired outcomes. The color of the utilized ink also influenced the accuracy of the outcome, given that the rate of conversion was different between various shades and colors of ink.