

HWD+ : A new Hand-written digit data set

Cédric Beaulac

October 31, 2020

General information

Files are saved in .npy format, these are Python numpy arrays.

Images(500x500).npy contains the full 13,580 images in 500 by 500 formats.

Images(28x28).npy contains the full 13,580 images in 28 by 28 formats.

WriterInfo.npy contains a matrix of size 13,350 by 9 containing Digit, ID, Age, Gender, Height(cm), Language, Hand, Education level and Writing Medium in that order.

Here is a descriptive data table:

Biological Gender	Male 46	Female 51		
Handiness	Right 84	Left 13		
Language (education)	French 75	English 16	Other 6	
Education Level	No high school 7	High school 13	Bachelor 55	Graduate 22
Usual writing medium	Hand 44	Keyboard 45	Other 8	

Table 1: Table of occurrence at the time of submission. The data set contains 97 writers.

More information can be found in the article presenting the data set : *Analysis of a high-resolution hand-written digits data set with writer characteristics*. The article is available on

the first authors website <https://cedricbeaulac.github.io/publications/>.

Code

Files can be imported with the code below :

```
import numpy as np
from numpy import *

Images = np.load(r"Images(500x500).npy")
WriterInfo = np.load(r"WriterInfo.npy")
```

Citation

If you are using this data set in any way, please cite the attached article :

Cédric Beaulac and Jeffrey S. Rosenthal, Analysis of a high-resolution hand-written digits data set with writer characteristics, preprint