Dating *Giacomo Leopardi*'s Texts (by Handwriting Recognition) using Convolutional Neural Networks

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

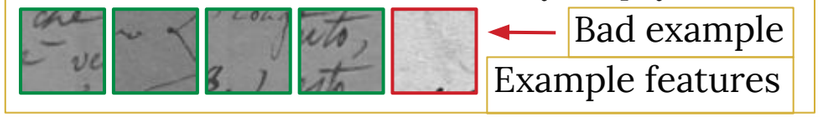
Pradyun & Nakul Solai, Aadhavan Raja MehtA+ Instructors: Bhagirath & Haripriya Mehta

Motivation

Classify Leopardi's Zibaldone and his other works by year (1817-20) using a Convolutional Neural Network.

Data & Preprocessing

1. Original manuscripts (organized into pages) were each dated sometime between 1817-20
2. We cropped these pages to remove areas that didn't contain text
3. Using OpenCV, a computer vision library, we grayscaled each of these images
4. To make our features, we cut these into small images of size 48x48, deleting the ones that had borders and were virtually empty



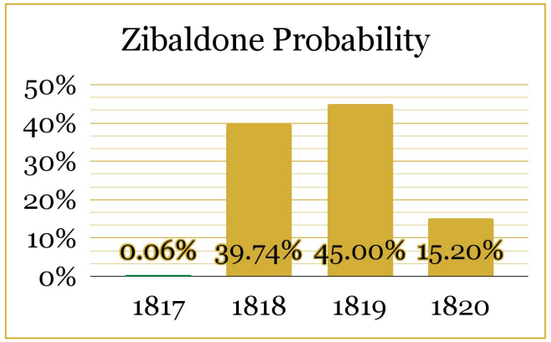
Bad example

Example features

Model

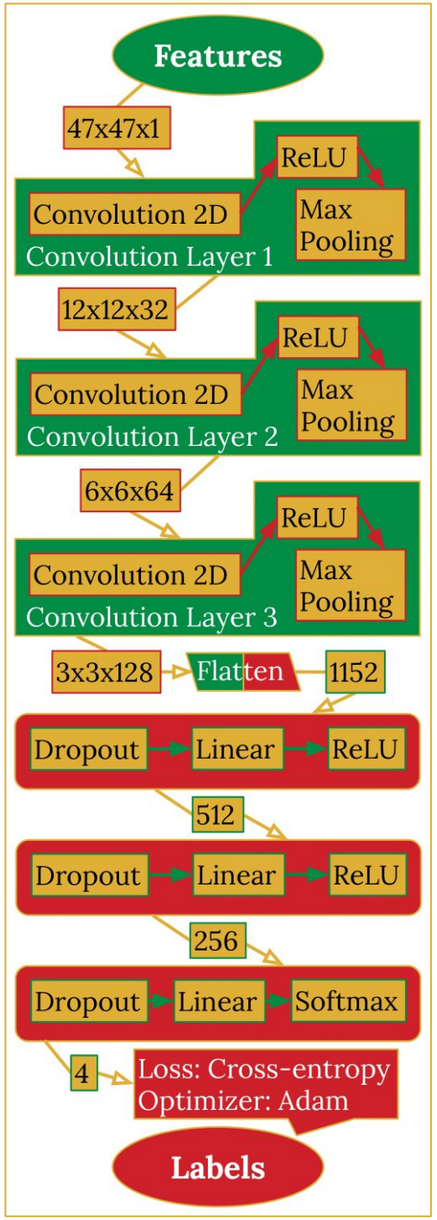
- Our Convolutional Neural Network (popular with image recognition) uses 60% of data for training , 20% for validation, 20% for testing
- Features go through 3 convolution layers, are flattened into 1 dimension of the same size, and go through 3 dropout and linear (dense) layers
- Categorical cross-entropy loss and Adam optimizer
- Structure is very similar to Dwivedi's [1]

Results



# of epochs	Epoch	Loss	**Accuracy**
100	100th (last)	0.5101	**85.20%**
100	70th (best)	0.3987	**87.28%**

- Labels: 0-3, represent 1817-1820
- converted into one-hot vectors
- Zibaldone manuscript isn't dated



Conclusion & Future Directions

- We predicted what years Leopardi wrote his texts and that the Zibaldone was most likely written in 1819
- Good that the model predicted 2 consecutive years (1818 & 1819)
- Use regression instead of categorical classification to predict exactly what day Leopardi wrote something.

References

[1] Dwivedi, Priya. "Handwriting Recognition Using Tensorflow and Keras." Medium, Towards Data Science, 24 Jan. 2018, towardsdatascience.com/handwriting-recognition-using-tensorflow-and-keras-819b36148fe5.