

Kaggle Competition: Find the “Largest” Digit

COMP 551-001, March 2018

Team Axolotl

Haohan Bo

260782229

haohan.bo@mail.mcgill.ca

Étienne Fortier-Dubois

260430244

etienne.fortier-dubois@mail.mcgill.ca

Martha Thomae Elias

260658140

martha.thomaeelias@mail.mcgill.ca

Abstract—The main text of the report should not exceed 6 pages. References and appendix can be in excess of the 6 pages. The format should be double-column, 10pt font, min. 1” margins.

I. INTRODUCTION

Briefly describe the problem and summarize your approach and results.

II. FEATURE DESIGN

Describe and justify your pre-processing methods, and how you designed and selected your features.

A. Pre-processing

Binarize images, isolate digits, find the largest one, remove the rest of the image, pad with black pixels and resize.

Isolate digits: one issue is that sometimes digits partially overlap. Separating these digits is a difficult problem. We have ignored it, because such cases are rare.

To determine the largest digit, we found the one that had the largest dimension, for either of the width and height. This approach proved superior to using the area (width \times height), as we found that images were classified according to the tallest or widest digit even though another digit may have had a larger total area.

III. ALGORITHMS

Give an overview of the learning algorithms used without going into too much detail in the class notes (e.g. SVM derivation, etc.), unless necessary to understand other details.

A. SVM or logistic regression

B. Neural network

C. Third approach

IV. METHODOLOGY

Include any decisions about training/validation split, distribution choice for naïve bayes, regularization strategy, any optimization tricks, setting hyper-parameters, etc.

V. RESULTS

Present a detailed analysis of your results, including graphs and tables as appropriate. This analysis should be broader than just the Kaggle result: include a short comparison of the most important hyperparameters and all 3 methods you implemented.

VI. DISCUSSION

Discuss the pros/cons of your approach & methodology and suggest areas of future work.

STATEMENT OF CONTRIBUTIONS

Briefly describe the contributions of each team member towards each of the components of the project (e.g. defining the problem, developing the methodology, coding the solution, performing the data analysis, writing the report, etc.) At the end of the Statement of Contributions, add the following statement:

We hereby state that all the work presented in this report is that of the authors.

REFERENCES

Optional.

APPENDIX

Optional. Here you can include additional results, more detail of the methods, etc.