Jan Čech
Center for Machine Perception
Department of Cybernetics, Faculty of Electrical Engineering
Czech Technical University in Prague

Bachelor's Thesis Review

Prague, January 14, 2021

Title: System for Automatic Checking of Solved Mathematical Equations in Raster Images
Author: Daria Tunina
Supervisor: doc. RNDr. Daniel Průša, Ph.D.
Date received: 08/01/2021

The thesis presents a method that processes scanned documents with simple mathematical equations meant for children (additions, or subtractions) with integer numbers 0–99. Left hand side is printed, while the results is handwritten. The algorithm detects and localizes all the printed equations and the handwritten numbers, recognizes them and checks for the correctness. The algorithm is implemented as a simple application with a graphical user interface.

The implementation is chained from popular libraries, e.g. YOLO and Tesseract, which is a standard engineering approach. YOLO detector was trained on a synthetic dataset prepared by the author of the thesis. I like that the implementation resulted in an application. I also appreciate the great effort to evaluate all components of the algorithm (localization, OCR for printed and handwritten parts) and the entire pipeline, together with an error categorization. Very interesting is an idea to estimate the final result from several photos.

On the other hand, the thesis suffers from several weaknesses. The major problem is that a level of details is very much imbalanced. The presentation mixes pure implementation details (such as input file formats or naming) with the algorithmic design of the method. Certain important details are revealed as late as in the chapter describing the application, other important details are not given:

1. The pipeline of the algorithm should have been presented as a flowchart or as a pseudocode. The text would be much more comprehensible.

2. The size of the synthetic dataset was not revealed. It would be worthy to mention how many different backgrounds, or fonts were used, what was the extent of the rotation, etc. How much of the extra text was inserted?

3. The handwritten part is expected to be up to 2-digit number. The thesis never describes, how two digits are recognized by the MNIST-CNN? Or, are the handwritten digits detected separately one-by-one by YOLO unlike the printed left-hand side of the equation? If so, how is the result composed? What happens if the digit bounding-boxes overlap, as e.g. in "69"? Did you consider any special treatment for this situation?

4. I expect, the image is rectified, so the content were axis-aligned. In Chapter 5, it is mentioned that this is done by a library function based on a paper border. How much reliable the rectification is? Did you consider estimating main directions from the inside text?

5. It is unclear, why the input images are Gaussian blurred prior to the recognition. Is it as an anti-aliasing filter prior to sub-sampling?

6. The thesis reports that accuracy can be improved by up to 10% if multiple photos of the same document is used. However, it is not detailed how it is done. How does the system combine several different outputs?

7. The entire pipeline is evaluated on the real dataset of scanned documents. A significant drop of accuracy is reported. How can you interpret the result? What was the accuracy of the pipeline on the synthetic test set?

8. A motivation of the work should have been explained better. The problem would be much easier if the printed document contains some machine readable markers, as e.g. QR codes. Then, locating the markers would solve the image rectification and localization of all objects.

Therefore, I suggest evaluating the thesis as

C – good.

Ing. Jan Čech, Ph.D.