Automatic Exam Grading by a Mobile Camera

Snap a Picture to Grade Your Tests

Benjamin Wagstaff
Electrical Engineering
UCLA
Los Angeles, California, USA
bwagstaff2020@ucla.edu

Chiao Lu
Computer Science
UCLA
Los Angeles, California, USA
josephlu1996@ucla.edu

Xiang ‘Anthony’ Chen
Electrical and Computer
Engineering
UCLA
Los Angeles, California, USA
xac@ucla.edu

ABSTRACT

Advances in mobile computing power have opened possibilities for processing intensive optical recognition and machine learning applications. Despite the widespread availability of mobile phones and digital resources, the education industry has made few updates to time-consuming grading techniques: many teachers continue to grade tests by hand or use outdated technology. Our goal is to produce a seamless, efficient, and accessible mobile experience to automate grading across multiple formats. We present a solution that allows teachers to take a photo and receive a visualized result after our deep learning application recognizes handwriting patterns, grades test answers, and identifies names and IDs. As a result, we contribute an improved computational tool for teachers to quickly create and grade exams across various formats.

CCS CONCEPTS

• Applied computing → Computer-assisted instruction • Human-centered computing → Mixed / augmented reality

KEYWORDS

machine learning; optical recognition; mobile grading

ACM Reference format:

Benjamin Wagstaff, Chiao Lu and Xiang ‘Anthony’ Chen. 2019. Automatic Exam Grading by a Mobile Camera: Snap a Picture to Grade Your Tests. In *24th International Conference on Intelligent User Interfaces (IUI ’19 Companion)*, March 17–20, Marina Del Rey, CA, USA. ACM, New York, NY, USA, 2 Pages. <https://doi.org/10.1145/3308557.3308661>

1 Introduction

In 2013, OKI Group published a paper, which found 95% of teachers still grade exams by hand [1]. Recognizing a need for exam grading solutions in the education industry.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

IUI ’19 Companion, March 17–20, 2019, Marina del Rey, CA, USA
© 2019 Association for Computing Machinery.
ACM ISBN 978-1-4503-6673-1/19/03...\$15.00
<https://doi.org/10.1145/3308557.3308661>

Grading exam papers is time-consuming. With the advance of OCR, OMR, and ML, tools that facilitate the grading process have emerged. In particular, Scantron was invented for grading bubble exams. However, competitive analysis shows these tools have drawbacks: they have unintuitive interfaces, they limit the types of problems teachers can ask, they lack convenient post-processing tools, and they are not universally affordable (requiring paid subscriptions or equipment). With this in mind, our goal is to build an easy-to-use tool to grade multiple exam formats that is free to use.

As an overview of our paper: we conducted a competitive market analysis, then surveyed teachers to analyze teaching uses and needs, and then we modularly designed our solution to address those unmet market needs.

1.1 Background and Related Work

Scantron has been widely used for exam grading since 1972. Since then, other exam grading apps have been introduced, to name some of the industry leaders: Zipgrade, GradeSmart, GradeCam, and PaperScorer. These apps use OCR/OMR to grade their respective proprietary bubble test formats using a phone or computer camera. However, they do not support other types of formats and do not recognize other test information. This causes additional processing time for teachers to identify names, ID’s, and test versions.

2 Survey Study

To understand how existing technology supports exam grading, we conducted a survey of 12 teachers across multiple areas of study from UCLA and Seattle Prep high school. Our hypothesis was that teachers spend too much time grading exams despite current solutions. From the survey we observed that 85% of teachers felt they spent too much time grading tests and 75% used multiple choice tests monthly (or more frequently). Additionally, of those 75%, most used Zipgrade or Scantron applications to grade class exams: noting it still took them upwards of 45 minutes to grade and record a classes’ worth of tests. In summary, the majority of the teachers believe that they spend too much time grading exam papers even with the help from current auto-grading technologies.

3 Snaptron

We introduce *Snaptron*, a mobile grading app, to help teachers efficiently grade exams. In short, a picture taken will be sent to a server for grading and the result will be presented back to the user's phone. We designed our system with two main objectives: (1) accessibility--letting users grade multiple exams from their phone and (2) robustness--allowing for user error in photo taking and paper recognition.

We implemented grading scripts for four different types of exam formats (our bubble style, Zipgrade style, one-word free response, and mixed) as self-contained demos.

Figure 1: (a) Snaptron User Interface, (b) paper detection, (c) backend architecture, (d) returned result.

3.1 User Display and Photo Taking

After opening the app, the user can select from four different types of formats to demo. The main screen displays photos taken using the scanning feature and their resulting annotated images. The user starts by clicking the camera icon at the top-right corner. To allow for amateur picture angles, *Snaptron* will identify the exam paper contours during live view, and after will let the user adjust the taken image (Fig. 1b). After submission, the user waits for about 5 seconds as the auto-grading script executes, and the graded version will show up on the screen; the user can click into the graded version for details.

3.2 Bubble/ZipGrade Format

Bubble exams, such as Scantron, are the most basic automatic exam format. In addition to grading bubbles, *Snaptron* can identify a student's ID and name from the Zipgrade format.

3.3 One-Word Free Response/Mix

Snaptron identifies written, one-word, answers filled into free response boxes and evaluates them against an answer key. Our mixed format is a combination of bubble, one-word, and true/false questions. *Snaptron* will grab all the bubble choice information, one-word responses, true/false questions, and the student's name/ID to completely grade a test paper.

3.4 Implementation: Front/Back End

We integrated the "WeScan" module [5] with our user interface to allow users to easily take, adjust, and upload scanned images (fig. 1b). Each image is uploaded and processed via the backend

servers and is then re-uploaded to the phone. Our backend consists of two parts: Firebase storage and an Azure processing server (fig. 1c). Firebase allows us to process each image as soon as its uploaded from a phone. The Azure server listens for incoming requests from Firebase, then responds by processing incoming documents using our Python modules. The graded result is returned to Firebase and presented on the mobile device in real time.

3.5 Grading Exams Using Computer Vision & Deep Learning

Using TensorFlow and Keras, we built two Convolutional Neural Networks (CNNs), both trained using the EMNIST dataset [4]. Each module takes an image array and outputs a matrix of probabilities. We built our models using several layers and activation functions: convolution layers, max pooling and average pooling layers, a locally connected layer, and several fully connected layers. The fully connected layers consist of: a dropout layer, a dense layer with ReLu activation, ending with a softmax regression layer. We used K-fold cross validation, an ADAM optimizer, and categorical cross entropy loss function to analyze our accuracy and results. Our handwritten number recognizer has 99.7% validated accuracy (after 6 iterations) and our handwritten letter recognizer had 94.4% validated accuracy (after 4 iterations). To compensate for the lower letter accuracy, we integrated a spell-checking Python library.

4 Conclusion

Snaptron is an education tool for exam grading using machine learning and OMR. Our survey and competitive analysis shows that there exists a need for a low-cost, fast, and robust alternative to traditional exam grading methods. Our implementation of *Snaptron* provides an example of a scalable product useful for not only grading multiple exam formats but also extracting important test information (such as name and student ID).

ACKNOWLEDGMENTS

We thank and acknowledge all the participants in our user surveys and reviewers for their time, as well as the HCI lab at UCLA for their support.

REFERENCES

- [1] Oki Data Americas, Inc. Bubble Sheet Testing in Education's Future. 2013. Retrieved December 5, 2018 from https://www.oki.com/us/printing/images/remark-test-grading-solution-white-paper_tcm75-143088.pdf
- [2] Adrian Rosebrock. PyImageSearch. 2016. Retrieved December 5, 2018 from <https://www.pyimagesearch.com/2016/10/03/bubble-sheet-multiple-choice-scanner-and-test-grader-using-omr-python-and-opencv/>
- [3] Adrian Rosebrock. PyImageSearch. 2015. Retrieved December 5, 2018 from <https://www.pyimagesearch.com/2015/04/20/sorting-contours-using-python-and-opencv/>
- [4] Cohen, G., Afshar, S., Tapson, J., & van Schaik, A. (2017). EMNIST: an extension of MNIST to handwritten letters. Retrieved December 5, 2018 from <http://arxiv.org/abs/1702.05373>
- [5] Avd Lee. GitHub. 2018. Retrieved December 5, 2018 from <https://github.com/WeTransfer/WeScan>