CS CAPSTONE PROBLEM STATEMENT

OCTOBER 17, 2017

REMOTE SEED IDENTIFICATION

PREPARED FOR

CROP AND SOIL SCIENCE DEPARTMENT, OSU

DAN CURRY

PREPARED BY

GROUP 12

QUANAH GREEN
ALEX RUEF
ETHAN TAKLA

Abstract

In a world where machine learning and computer vision are exponentially growing industries, combing through thousands of seeds to find rouge ones seems quite archaic. Currently, individuals at the Oregon State Seed lab sift through thousands of seeds a day using nothing more than some simple machinery and microscopes. Seed analysts in the field have to send their samples to expensive testing facilities that take days to process. The seed identification business isn't too big of an industry, so a high-tech solution for seed processing hasn't been developed yet. A mobile, robust, and efficient solution to these problems is proposed in this paper, in which clients can send images of seeds via their mobile phones to be processed, and get the results in minutes.

CONTENTS

1 DESCRIPTION

Seed identification at Oregon State is a time-consuming, inefcient process that often leaves analysts with eyestrain and headache. A seed analyst will look over up to 25,000 seeds a day, which is currently done using a vibrating conveyor belt apparatus with magnifying lenses for the human operator. In addition to in-lab testing, seed analysts often have to take samples in the field and mail the seeds to remote locations that do the analysis for them. This is also a long, expensive process that only contributes to the urgency of finding a more elegant solution for seed identification. One of the primary reasons that nobody has developed a solution for this yet is that grass seeds are extremely small and diffcult to differentiate from one another without using expensive equipment. Instead of just being able to determine whether there is a bad seed in the sample, it would also be advantageous for a solution to this problem to correctly classify what type of seeds are in the sample. A typical sample consists of 2500 seeds, meaning identification with a normal resolution camera can be difficult.

2 SOLUTION

In light of the glaring inefficiencies and issues with Oregon State's current seed analysis methods, our team proposes a robust, mobile solution that can be used in the field, and doesn't require expensive camera equipment. Using an IOS or Android based smart phone or tablet, analysts can take pictures of the seed samples, which are sent to a remote NVIDIA Jetson TX2 processor that runs machine learning algorithms to correctly identify rogue seeds. Once the processing is complete, the results are sent back to the client, speeding up the seed identification process enormously. Each seed sample will be contained in a pre-made 3D-Printed tray with dimension markers on the edges to improve classification of different-sized seeds. For image classification, both NVIDIA and OpenCV have powerful, multi-threaded, deep neural network libraries that can run on the Jetson's 256 CUDA cores for quick results. Upon the return of a result, the client will be able to see where the abnormal seeds are in the sample, and what type of seed they are. On top of all this, we hope to allow an easy way for analysts to add new types of seeds to the image classifier, so that the system can stay current without the need of computer science specialists constantly updating the image classifiers.

3 PERFORMANCE METRICS

Upon completion of the project, a user-friendly, portable, and efficient seed identification solution will be in the hands of OSU seed analysts. Due to the novel nature of this project, there is a degree of uncertainty inherent within whether or not our ultimate ambitions for the end product will be met. In order to consider the project completed, the bare-minimum following components must be completed: 3D printed seed tray, intuitive client application, web server for communication with the Jetson TX2, image classifiers with a greater than 95 percent positive yes/no seed identification rate (for a 500 seed sample), and an easy to understand image and data output that gives the analysts important metrics on the sample. The yes/no image classifier can tell the client which seeds aren't like the others, but cannot properly identify what species the rogue seeds actually are. Ultimately, the goals is to have a classifier that can identify the correct species of each of the seeds as well. Another ambitious goal set for this project is to be able to use 2500 seed samples, rather than just 500, to speed up the process.