CSCI 1430 Final Project Report: Analyzing Leaf Morphology

Analyzing leaf morphology: Henry Dawson, Daniel Healey, Cheyenne Chau.
Brown University

Abstract

For our final project, we created a model to classify images of plant cells based on image quality. Our approach was to use a convolutional neural network. This network was trained on a dataset containing good quality cell images and their respective artificially degraded counterparts. The quality of images in the training and testing dataset was labeled by project participants on a subjective basis. The convolutional neural network preformed well on the training set of good and poor quality images.

1. Introduction

Research in the fields of biology and medicine often involves on the manipulation of images of cells, leading to many applications of computer vision in cell research. However, many high-level computer vision tasks, such as cell segmentation, rely on good quality image data both during training and deployment. Models for these high-level tasks do not extrapolate well to images that are out of focus, noisy, have poor illumination, have blur due to motion, or contain obstructions.

Poor quality images occur due to a myriad of factors. Human error, equipment issues, and environmental factors all have an effect on the quality of cell images taken by a microscope. While these factors can be mitigated, the possibility of poor quality images will remain. Therefore, it is often necessary to incorporate detection of poor quality images as one of the first steps to computer vision pipelines.

While these images can be detected manually, automating this process with computer vision would be more efficient. Thus, for our project we attempted to create a model to detect poor quality images.

2. Related Work

Previous work on image classification of microscopic images has been done by Kose et al. [1], Tennakoon et al. [2], and Yang et al. [3].

In their research, Kose et al. [1] investigated the application of computer vision for detecting image quality of lesions meant for in vivo reflectance confocal microscopy. Due to the fact that analysis of these images would be done by a doctor at a later time, possibly after the patient was no longer easily available for subsequent images, it is important to assess the quality of images taken of these lesions. While their approach involved segmenting the image based on informative and uninformative regions of the image, their research was still applicable to ours as it shows the feasibility of computer vision for image quality detection for research in the fields of biology and medicine.

Convolutional neural networks in particular have seen increased use for image quality classification. Tennakoon et al. [2] applied a convolutional neural network for image quality assurance in a automated diabetic retinopathy screening system. They used a pretrained deep convolutional neural network to screen retinal images based on quality, achieving high accuracy. Due to the effectiveness and ease of implementation of convolutional neural networks, we decided to use a convolutional neural network as our model.

Much of our paper was based on the work of Yang et al. [3]. In their work, Yang et al. applied a shallow convolutional neural network in order to classify images of cells based on focus. However, instead of training on just in-focus and out of focus images, the researchers trained on in-focus and artificially defocused images. These images were defocused by convolving the image with a point spread function defined by the equation in Figure 1.

$$h(x,y,z) = \left| C \int_0^1 J_0\left(k\frac{NA}{n}\sqrt{x^2+y^2}\rho\right) exp\left(-\frac{1}{2}jk\rho^2 z\left(\frac{NA}{n}\right)^2\right)\rho d\rho \right|^2$$

Figure 1. Defocusing equation

By training their model on artificially defocused data, Yang et al. [3] was able to increase the effectiveness of their model without the requirement for more data. Because of our limited data, we decided to take this approach, opting to train on original and defocused good quality images, and saving

the limited poor quality images that we had exclusively for testing.

3. Method

The approach that we took was to train a convolutional neural network to detect image quality. The model's architecture was based heavily on that of Yang et al. [3], with two convolutional layers consisting of thirty two 5x5 filters each, two max-pooling layers, and three fully connected layers.

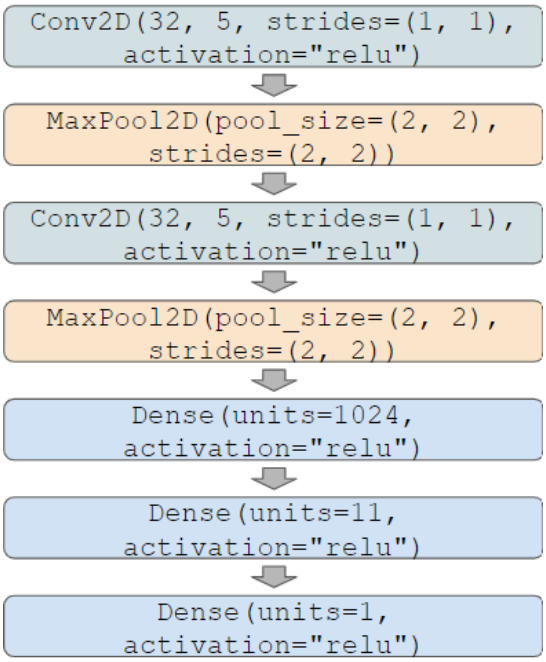


Figure 2. Model architecture

This model was trained and tested on image data of epidermal plant cells. These plant cell images were sourced by one of the project members. While many of the images were taken by the project member, some were taken by others and supplied to him for use in the project. Theses images were unlabelled, so labelling was preformed by project members. The images were labelled with binary good/poor labels for image quality. These labels were assigned on a subjective basis, but were based primarily on the image's focus, blur and any obstructions in the image.

The image data was split into two parts: training and testing. The training data contained good quality images as well as degraded versions of these images. While the majority of good quality images in the training dataset had a corresponding degraded copy, not all did. A split of the training data was preformed, such that 20% of the training data was used for validation and the other 80% was used for training the model. The testing data contained only unadulterated images, some of which were poor quality and some of which were good quality. The testing data did not contain any artificially degraded images. There was a roughly equal split of good/degraded images in the training data set and a roughly equal split of good/poor quality images in the testing datset. A summary of the data can be seen in Table 1.

Partition	Elements
Training	627
Validation	156
Testing	31

Table 1. Image data summary

During training, data augmentation was preformed, in order to increase the effectiveness of the resulting model. Both the training and testing data was resized to be of size 256x256, and then converted to grayscale. Training images were then rotated by up to 30 degrees, flipped horizontally and flipped vertically. This helped to prevent our model from overfitting to our somewhat limited dataset. These data augmentations were not preformed on the testing data.

After we had settled on an architecture to use for our convolutional neural network and a method to train it, we preformed a study on how certain hyperparameters and data augmentation during training influenced the performance of our model. For hyperparameters, we investigated how batch size and number of epochs improved our performance. For data augmentation, we investigated zoom range, shear range, samplewise centering, samplwise normalization, horizontal flipping and vertical flipping.

4. Results

Before training or testing of our model could begin, we had to source our image data. Data were collected as part of lab work done by a group member. The training data set was composed of images at multiple magnifications and of varying quality and content, within the same broad category. Specifically, all images were of the epidermis of Arabidopsis thaliana, imaged using brightfield microscopy. Images were prepared by growing and harvesting plant material. Leaves were pressed gently to a slide covered in partly cured adhesive. After drying and being re-moistened, leaves were softly scraped using a pipette tip as to remove all but the most superficial epidermal cell tissue. The remnant tissue was stained briefly to aid visualization of cellular material and then mounted. Images were taken on an Axioplan 200M. These images were then degraded using the defocusing equation (Figure 1), which mimics the optical effects of a defocused microscope image.

Training then worked to classify images as either quality or degraded based on focus. The model ran until early stopping halted training when a sufficient level of accuracy had been stabilized. We assayed model performance and training by varying batch size, loss, and the optimizers (Table 2).

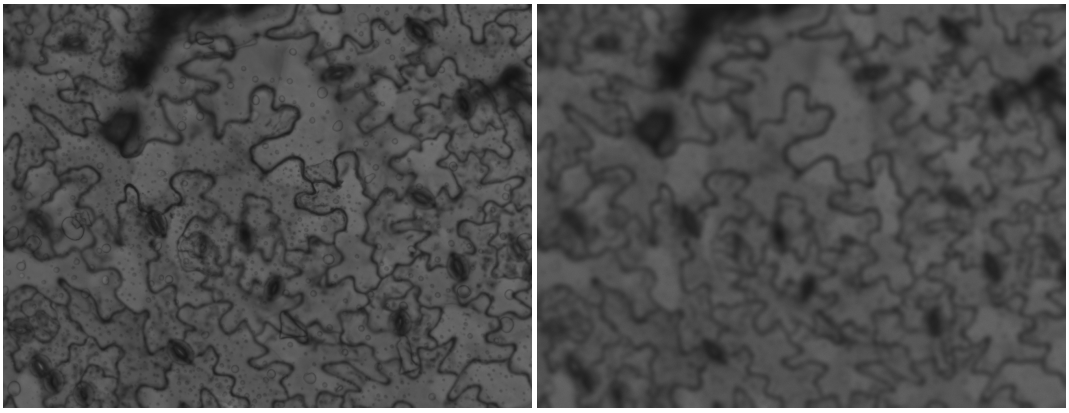


Figure 3. Image degredation. *Left:* Original image. *Right:* Degraded image.

Batch Size	Loss	Optimizer
8	Binary Crossentropy	Adam
16	Categorical Crossentropy	Adagrad
32		Adadelta
		sgd

Table 2. Parameters tested in ablation studies

In general, the model trained to a high level of accuracy – roughly 96% – in a relatively short time frame (Figure 4), between 2 and 20 epochs. This mean that early stopping could be used in order to prevent use of our limited compute power for unnecessary training.

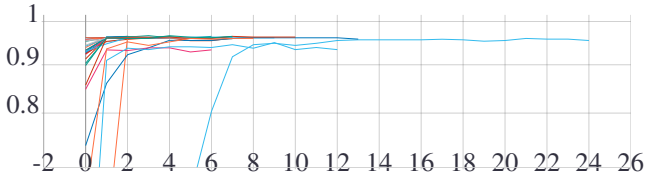


Figure 4. Summary of training runs (batch size 32)

Varying optimizers and loss functions resulted in minimal functional improvements in general (Tables 3 and 4). All optimizers plateaued to similar accuracy within the same general timeframe. Of the loss functions, categorical crossentropy minimized epochs needed to reach the early stopping criteria, whereas binary CE stopped later. Trends in validation accuracy closely follow those shown here, which are from training.

4.1. Technical Discussion

Our implementation was successful but ultimately lightweight. Specifically, where the paper classified images into 10 qualities of focus, ranging from best to worse, we operated within a binary due to limitations in data quality brought on by a focus in data acquisition of taking in-focus images. In addition, the original paper featured further architecture to segment an image into a grid of patches and assign each a quality score, which we did not implement. These concessions were due to data availability and the scope of our project, but would be interesting components of implementation. In addition, only having two distinct classes, with a clear boundary – the blur – between the two facilitated fast and reliable training, which otherwise may have taken more fine tuning. It is likely that the facility of training would decrease with added classes.

4.2. Societal Discussion

Although our work will have a net positive societal impact, we would like to mention that certain groups may experience negative effects. Foremost of these groups are the people who rely on manual image labeling as their primary source of income. For these people, advances in technology would threaten their jobs by decreasing demand.

However, overall our project will have a net positive societal impact. By increasing the efficiency of high-level computer vision tasks used in biology and medicine research, society will reap the benefits of this research. Advances in biology and medicine will improve agriculture and medical care internationally.

5. Conclusion

We implemented a model architecture to classify microscopic images as in-focus or out of focus, a general determinant of image quality. We applied an optics-based degradation function to generate training data from biological research data, and used this data to succesfully train our model. We conducted ablation studies to assay the impacts of various model choices on our model, and ultimately concluded that training on our dataset was viable with any combination of parameters tested, with negligible differences in training time.

Massive amounts of modern biological research hinge on image analysis. At the fundamental level, interactions with

light provide the highest-resolution data for biological specimens. These images are often hand-annotated, a practice that is efficient for ensuring accuracy but introduces potential bias, labor issues, and massively limts throughput. Development of tools to facilitate automated image analysis are poised to greatly impact biological research by minimizing the amount of manual labor needed for many phenotypic experiments and maximizing the usage of model organisms, a key factor in many labs. Quality control is important as a standalone aid for image processing; it can be used to limit a researcher from anayzing poor quality images. It can also play a key role in development of full automated analysis pipelines; images can be taken and filtered according to quality, preventing a model from operating on sub-par images.

Moving forward, models like these need to be developed holistically into the research process. Their incorporation into experimental design necessitates a variety of considerations; that training data be de-biased (e.g. using multiple genotypes, not just wild-type), that accuracy thresholds be determined beforehand, that scientists maintain the ability to meaningfully interface with the outputs of their code and make sense of the models predictions. Adherence to scientific principles while implementing cutting edge computer vision techniques will aid seamless translation.

References

[1] Kivanc Kose, Alican Bozkurt, Christi Alessi-Fox, Dana H. Brooks, Jennifer G. Dy, Milind Rajadhyaksha, and Melissa Gill. Utilizing Machine Learning for Image Quality Assessment for Reflectance Confocal Microscopy. *Journal of Investigative Dermatology*, 140(6):1214–1222, June 2020. 1

[2] Ruwan Tennakoon, Dwarikanath Mahapatra, Pallab Roy, Suman Sedai, and Rahil Garnavi. Image Quality Classification for DR Screening Using Convolutional Neural Networks. In *Proceedings of the Ophthalmic Medical Image Analysis Third International Workshop*, pages 113–120, Athens, Greece, October 2016. University of Iowa. 1

[3] Samuel J. Yang, Marc Berndl, D. Michael Ando, Mariya Barch, Arunachalam Narayanaswamy, Eric Christiansen, Stephan Hoyer, Chris Roat, Jane Hung, Curtis T. Rueden, Asim Shankar, Steven Finkbeiner, and Philip Nelson. Assessing microscope image focus quality with deep learning. *BMC Bioinformatics*, 19(1):77, March 2018. 1, 2

Appendix

Team contributions

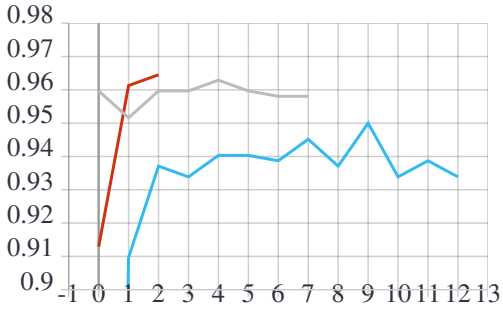
Please describe in one paragraph per team member what each of you contributed to the project.

Henry Dawson Worked on initial paper proposal, researching related works, sourcing all image data, labeling the training/validation data, generating degraded images for training/validation, coding the model/runner, fine tuning hyperparameters, and final paper.

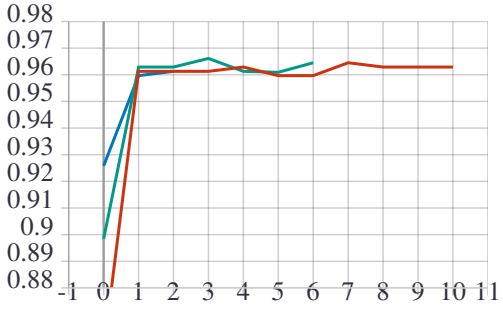
Daniel Healey Worked on initial paper proposal, researching related works, labelling of the testing data set, coding the model/runner, and final paper.

Cheyenne Chau Worked on the initial paper proposal, final paper, created final poster, and helped with the socially responsible aspects of the project.

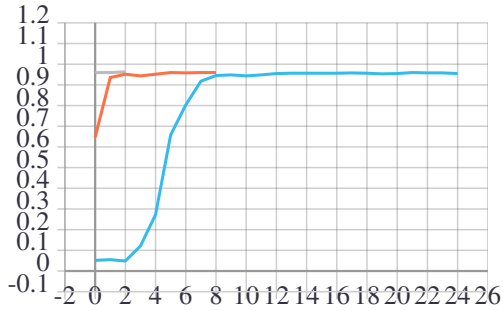
Table 3. Optimizer variation in training (batch size 32), Accuracy vs. Epoch. Different lines represent varied loss functions and batch sizes.



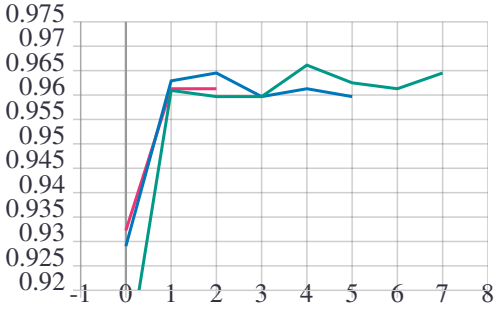
Adam



Adagrad

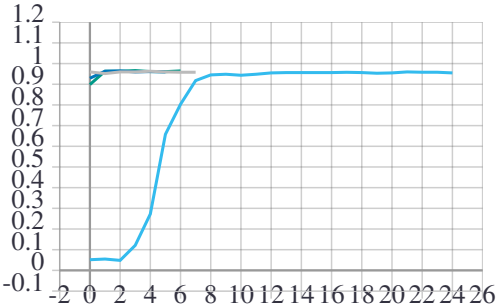


Adadelta

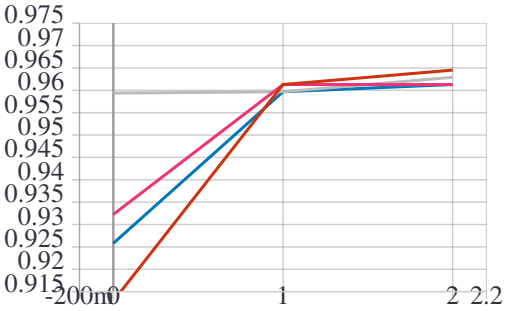


SGD

Table 4. Loss variation in training (batch size 32), Accuracy vs. Epoch. Different lines represent varied optimizers and batch sizes.



Binary CE



Categorical CE