Team name: Joker
Team members: Jiacheng Zhou, Ruitao Shen, Shihang Liu, Siwei Guo, Yuanchi Guo
Date: 04/12/2024

This journal has to be uploaded to 1) Canvas (will create assignments)**. AND** 2) posted to your project website every two weeks by 11:59 pm on the second Friday (except the spring break). When uploading to Canvas, change the file name to "Team_name_MM-DD-2023" where MM = month and DD= day of upload).

Team roles for this report (write down name):

Facilitator(s): Siwei Guo

Recorder(s): Ruitao Shen, Shihang Liu

Deliverer(s): Yuanchi Guo

Planner(s): Jiacheng Zhou

See last page for description of roles. Obviously one person can take more than one role or there can be more than one person per role or make your own roles!

0. **Describe briefly what the main goal of your team is (so the peer reviewer has some context). E.g. we are working on image classification for blah de blah. Our goal is blah de blah etc. In the initial part of the semester before your proposal it is ok to put down "we are still coming up with ideas on team project".**

Our project is centered on image classification, and after careful consideration, we have decided to focus on fruit and vegetable classification rather than animals. This decision was made due to the significant variability within animal species, which could potentially complicate the identification process. Our specific objective is to employ supervised learning techniques to train a model capable of distinguishing between similar fruit categories, such as different types of apples, based on subtle physical features like colors and

shapes. This classification is crucial as similar fruits often lead to confusion among consumers, sellers, and market managers. While our current emphasis is on fruit, our direction may evolve as we progress. Our group's plan includes experimenting with dimension reduction techniques, specifically comparing t-SNE with PCA to determine which yields better results. We also intend to implement coding for Convolutional Neural Networks (CNN) and random forest algorithms using a small training dataset. Additionally, we are exploring new, reliable, and efficient dimension reduction algorithms that are applicable to our project. We are committed to coding these algorithms and integrating them with our training data to enhance the accuracy and efficiency of our classification model.

I. **What was done during the report period regarding the project:** If you want to include code include this in the Appendix. Describe what the group did (including contributions of individual team members) with regards to the group project during this report period. Give enough details so I understand what you folks have been doing over the week. Include dates of your meeting(s) and who met on these days.

Our group consistently meets every Sunday morning from 10 am to 11 am. During the meeting on 3/31, we discussed strategies for achieving dimension reduction on the entire training dataset, as current algorithms have only proven efficient with limited training data and struggle when applied to the entire set. Jiacheng and Yuanchi mentioned about the hyperparameter tuning method to find the best set of hyperparameters for a given algorithm and dataset. Shihang and Ruitao suggested the idea of implementing PCA by combining R Studio and Python to achieve flexibility and different analytical tools through dual-language applications. Siwei experimented with Colab using Python and R, instead of relying solely on local R Studio, to implement PCA.

In the meeting on 4/7, we explored additional dimension reduction algorithms identified through literature review and discussed the results of our CNN experiments. Half of the members shared opinions on the CNN results, while the other half presented findings related to dimension reduction algorithms. Jiacheng and Yuanchi said after compiling the model with an optimizer, loss function , and metrics, they configured Image Augmentation and fitted the Model. Shihang and Ruitao shared the idea of locally linear embedding (LLE), which is a non-linear method of dimension reduction that is excellent at preserving the local structure of the data. Siwei shared insights on using t-SNE as our dimension reduction technique since t-SNE performed more efficiently than PCA when applied to the expanded training dataset and also suggested on transitioning from R to Python as our primary coding tool.

II. **What were obstacles faced if any in working on the project?** This could be technical (like not being able to implement or understand particular techniques) or time issues (midterms for other courses etc).

Currently, we are encountering a specific obstacle related to coding and execution. When attempting to apply the PCA code to the entire image folder, RStudio initiates the process but fails to complete it. However, if we

include too few images in the training sets and subsequently apply the model to test sets with fewer images, the credibility of the model and its performance will be compromised. While we have t-SNE as one backup plan and are considering alternative coding tools, we are unsure if these options will yield significant benefits. But the evaluation of the model does not seems to be so good so we are still adjusting details of our model performance to make a more robust detection model to best utilize the data.

III. What is the plan for the next reporting period including what each team member is planning to work on. Describe goals and potential timelines (" I plan to finish understanding x to see if it can be implemented for our project by Wednesday etc".)

Before the next reporting period, our objective is to fully complete dimension reduction using either t-SNE on any coding platform, and to finalize the coding for three classification algorithms to be included in our final project. Siwei will endeavor to expand any dimension reduction methods implemented on Colab/Python/R to apply to all images, rather than just a small subset, and will also strive to finalize the code for random forest. Shihang and Ruitao will continue on implementing KNN to the project and finalize the coding part of CNNs based on the result of our CNNs experiment. Jiacheng and Yuanchi will develop a random forest classifier as a comparative model, train both models on the training set and validate using the validation set.

Train both models on the training set and validate using the validation set.

While in the biweekly document above you will describe what your team did with regards to the team project (with proper attributions of who did what in the week) there are 4 pre-defined roles. I urge you to have different people do these jobs every week so that you gain experience in each of the jobs. There can also be more than one person per job for example 2 people recording the weekly journal.

Facilitator: Manages the group for this week including setting up times for group members to meet, making sure everyone has a say in the meetings etc.

Recorder: Person in charge of recording the meetings as well as the happenings of the past two weeks and describing what was accomplished in the meeting and writing up this report.

Deliverer: Person in charge of checking the entire report and uploading the file to dropbox folder and project website, as well as the representative of the group getting in touch with the instructor.

Planner: Person in charge of what will be happening next two weeks as well as thinking about longer term goals (what more needs to be done for the project).

Team contact: Person I can email if I see any issues in the biweekly report instead of mass spamming everyone in the team.