Team name: Joker
Team members: Jiacheng Zhou, Ruitao Shen, Shihang Liu, Siwei Guo, Yuanchi Guo
Date: 03/29/2024

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 focuses on image classification, with a decision to pursue fruit and vegetable classification over animals due to the significant variability within animal species, which could complicate identification. Our specific goal is to employ supervised learning to train a model capable of discerning between similar fruit categories, such as different types of apples, based on subtle physical features like colors and shapes. This classification is essential as similar fruits often confuse consumers, sellers, and market managers. While fruit

is our current focus, our direction may evolve as we progress. Our group plan includes ongoing preparation of more images for analysis, utilizing PCA for dimension reduction, and experimenting with code for CNN and random forest on a small training dataset. Additionally, we are exploring new, reliable algorithms applicable to our project and endeavoring to code them with our training data.

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. In the meeting on 3/10, we deliberated on specific coding about dimension reduction and continued self-studying over interesting machine learning algorithms for image classification to include in our project. Jiacheng and Yuanchi re-examine photo formats, gallery diversity and tags based on the last CNN model. Shihang and Ruitao continued doing research about some example codes for the final model and began building the final research model using CNNs. Siwei continued solving coding errors and problems using PCA, gradually expanding its application to a larger training dataset.

In the meeting on 3/17, we discussed feedback from Professor Zhang on our research proposal. Each member shared opinions on how to improve our project, especially regarding algorithm choice and method application. Under the suggestion, Jiacheng and Yuanchi re-examined the knn method, but it is very time-consuming and is not hirecahical feature representation, so the feature extraction process is very complex and time-consuming. Then we use their models to process the initial image library and see what the effect is. Ruitao and Shihang shared the idea about KNN, which is a algorithm that is used for multi-class classification and make classification of a new image based on the images we have in our dataset. Siwei learned and shared about using Random Forest as a learning algorithm, which operates by constructing multiple decision trees during training and outputs the mode for classification of the individual trees and has the ability to handle large datasets with high dimensionality, noise, or missing data effectively.

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 facing a specific coding and running obstacle. When applying the PCA code to the entire image folder, RStudio runs but never finishes. However, if we include too few images in the training sets and apply the model to test sets with fewer images, the credibility of the model and its performance will be reduced. We are still working on solving specific coding errors for random forest and CNN.

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, we aim to finalize approximately three classification algorithms to apply in our project and nearly complete the coding part for dimension reduction and models. Siwei will strive to extend PCA to apply to all images rather than just a small subset and attempt to complete the code for random forest. Shihang and Ruitao will try to finish the coding part of the final research model using CNNs and understanding KNN to see if it can be implemented for our project. Jiacheng and Yuanchi focus on reshape image size and also use augmentation to apply transformations for better performance overall and normalization to scale pixel values.

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.