# Research background

What I will go first is the research background.

The reasons why we choose the task of fruits classification are as follows:

On one side, fruits have certain categories that are hard to differentiate, like the citrus genus, that contains oranges and grapefruits. Thus we want to see how well can an AI model complete the task of classifying them.

Another reason is that fruits are very often found in stores, so they serve as a good starting point for building more complex projects.

What’s more, researching a successful fruit classification model is of great importance.

Fruits, as an important component of human daily diet, also play a crucial role in the food industry

Besides, by solving fruit classification, we can further apply it to many situations such as automated fruit sorting, supermarket automatic replenishment, and autonomous fruit harvesting.

# Challenges of Fruit Classification

When conducting research on the Fruit 360 classification project, we also noticed a series of problems and challenges that the project is currently facing. I will highlight a few of them.

Firstly, we have observed that the project is affected by external factors. For instance, the lighting environment is quite complex, leading to shadows and reflections on the fruits. This poses a challenge for the model to accurately classify the fruits. Additionally, when there are multiple fruits present in a single image, it becomes difficult for the model to detect and classify each type effectively.

Secondly, the diversity of the dataset is a major concern. We have noticed an imbalance in the variety of fruits included in the dataset. This can result in overfitting issues when working with small samples and may cause the model to be biased towards recognizing common fruit categories while ignoring rare ones. Furthermore, variations in fruit ripeness and appearance further complicate the classification process, as the model tends to focus on recognizing common categories rather than rare ones.

Although the challenges are many, we will try our best to solve them

# SVM

In the research, we decided to use several models to classify the fruits. I will show you tow of them.

The first one is SVM.

Support Vector Machine (SVM) is a commonly used supervised learning algorithm for binary and multi-class classification tasks. It is often chosen for its efficient classification in high-dimensional space and its ability to perform well on small samples. This makes SVM particularly suitable for fruit classification, where multiple features such as color and size need to be considered.

The core idea of SVM is to find an optimal hyperplane that can separate different classes of samples. The objective is to maximize the margin, which is the distance between the hyperplane and the support vectors - the sample points closest to the hyperplane. By maximizing the margin, SVM enhances the robustness and generalization capability of the classification.

SVM employs different techniques for different data scenarios. For linearly separable data, it uses hard margin maximization to find the optimal hyperplane. However, when faced with linearly inseparable or noisy data, SVM introduces the concepts of soft margin maximization and kernel functions.

The ability of SVM to adapt to different data distributions is another advantage. By selecting appropriate kernel functions, SVM can effectively handle various data patterns.

Considering these advantages, we have decided to incorporate SVM as one of our fruit classification models.

# KNN

直接照着ppt念

# About data set

Describe the data set in general. It collects a large number of fruits, such as apples, strawberries, and so on, and contains fruits of the same kind but different shapes.

There are some properties of the data set.

Total number of images: 90,483.( Ninety thousand, four hundred and eighty-three.)

Training set size: 67,692 (Sixty-seven thousand, six hundred and ninety-two) images

Test set size: 22,688 (Twenty-two thousand, six hundred and eighty-eight) images (one fruit or vegetable per image).

Number of species: 131 (fruits and vegetables).

Image size: 100x100 pixels (One hundred by one hundred pixels).

In addition, I would like to explain how these images were obtained by planting fruits and vegetables on the axis of a low-speed motor (3 rpm), using a webcam and recording a 20-second short film.

Behind the fruit, the author of the dataset put a white piece of paper as a background.

Then due to changes in lighting conditions, the background is not uniform, and the author used an algorithm to extract the fruit from the background. The algorithm is the flood fill type: start at each edge of the image, label all pixels there, and then label all pixels found near the labeled pixels whose colors are less than a specified value apart. And repeat the previous step until no more pixels can be tagged.

In general, all marked pixels are treated as background (then filled with white), and the remaining pixels are treated as belonging to objects.