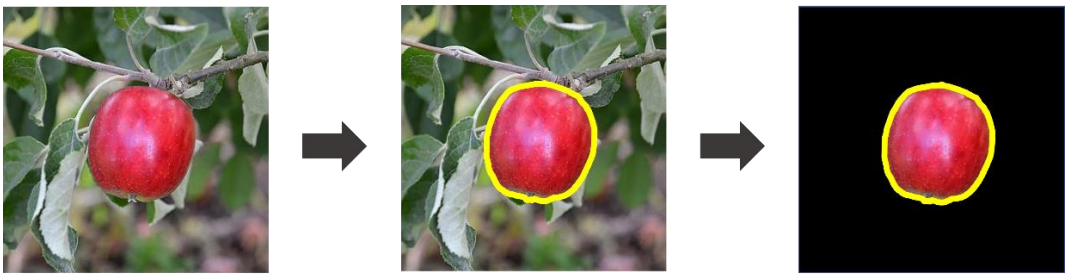
# Lab 11
# Machine Learning II

## A. Feature Extraction and Classification (50 points)

In this section, you will focus on extracting meaningful features from a dataset of fruit images, which will lay the foundation for subsequent machine learning classification. Initially, you will extract the positions of fruits in the images, distinguishing them from the background to obtain clear information about their location and shape. This process is crucial as it directly influences the effectiveness of the machine learning models you will later apply. The goal is to understand how accurately and effectively different features can be extracted and how they impact the performance of various machine learning algorithms.

1. Extract the contours of fruits from the images in the fruit dataset. Fill in the contours with a mask to achieve separation of the foreground and background. Refer to the provided example image for guidance. (10 points)



(a) (b) (c)

Figure 1. Process of fruit extraction.

2. Use Fourier Descriptors to transform the contours extracted in the first section into high-dimensional vectors of various levels. (10 points)
3. Calculate the texture features of the extracted fruits (Figure 1.c) using the Gray-Level Co-Occurrence Matrix (GLCM) practiced in the course. (10 points)
4. Organize and compile the extracted features into a data frame as illustrated in the provided figure. (10 points)

| | filepath | class | label | phase | dissimilarity | homogeneity | contrast | energy | ASM | correlation |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | Fruitsv2\train\Apple\-1x-1_jpg.rf.e470eabbf1e6... | Apple | 0 | train | 4.433679 | 0.479994 | 95.366361 | 0.068094 | 0.004637 | 0.965118 |
| 1 | Fruitsv2\train\Apple\006b4e6f706e3119_jpg.rf.3... | Apple | 0 | train | 2.757084 | 0.675332 | 152.035987 | 0.271609 | 0.073771 | 0.965764 |
| 2 | Fruitsv2\train\Apple\00_m_jpg.rf.cac94222ac6af... | Apple | 0 | train | 6.487515 | 0.296899 | 220.685748 | 0.037494 | 0.001406 | 0.956335 |
| 3 | Fruitsv2\train\Apple\0148_jpg.rf.7f31c6a9270b4... | Apple | 0 | train | 9.191914 | 0.156408 | 237.483494 | 0.018561 | 0.000345 | 0.857710 |
| 4 | Fruitsv2\train\Apple\015415912756_jpg.rf.fdb37... | Apple | 0 | train | 10.391545 | 0.130731 | 282.260551 | 0.012375 | 0.000153 | 0.939261 |

Figure 2. Data frame.

5. Use the three machine learning models introduced in the course (Decision Tree, K-Nearest Neighbors (K-NN), Support Vector Machine (SVM)) to classify the features described above. (10 points)

## B. Discussion (50 points)

In the first part, you extracted the positions of fruits in the images, removed background noise, obtained information about the location and shape of the fruits, and used this for training a machine learning model. In this section, you will explore the impact of the machine learning model and training data on the model's performance.

1. Answer the questions below (10 points)
    1) What is the highest validation accuracy you achieved?
    2) Explain the machine learning model approach you used.
2. Answer the questions below (20 points)
    1) The features used for training the machine learning model can influence the final accuracy. Compare the effects of adding and removing different features on the model's accuracy and explain their possible reasons.
    2) In your opinion, how many orders of Fourier descriptors can achieve the highest accuracy for the model?
    3) Does removing the background enhance the model's accuracy, and how effective is this enhancement?
    4) Which features did you use to train the machine learning model?
3. Answer the questions below (20 points)
    1) Identify images where the model misclassified objects and explain possible reasons for the misclassification.
    2) Propose solutions to prevent these misclassifications (provide ideas only, no need for implementation).