

# RipeTrack: Assessing Fruit Ripeness and Remaining Lifetime Using Smartphones

Muhammad Shahzaib Waseem, Neha Sharma , and Mohamed Hefeeda , *Senior Member, IEEE*

**Abstract**—Several studies have shown that a significant fraction of fresh fruits is discarded at the retail and consumer levels, wasting precious resources, polluting the environment, and contributing to increased food prices. An important factor contributing to this problem is the lack of scalable solutions for determining fruit ripeness and remaining lifetime. We propose a cost-effective solution that leverages the sensing capabilities of phones and machine learning models to analyze the optical properties of fruits at various ripening stages. The proposed solution is non-invasive, works for different fruits, and produces intuitive outputs, e.g., Unripe/Ripe/Expired and the percentage of remaining lifetime, enabling retailers and consumers to minimize food waste. We implement a proof-of-concept mobile application, RipeTrack, and demonstrate the accuracy and robustness of the proposed approach using an extensive empirical study with multiple fruits, including avocados, pears, bananas, nectarines, and mangoes. Our results show, for example, that RipeTrack can identify the ripeness level of avocados and pears with an accuracy of 95% and 98%, respectively, and it can predict their remaining lifetimes with an accuracy of 93% and 97%. Our results also show that RipeTrack can easily be extended to new fruits using transfer learning, and it functions in realistic environments, e.g., homes and grocery stores, that have diverse illuminations.

**Index Terms**—Fruit ripening, mobile applications, hyperspectral imaging, spectral analysis.

## I. INTRODUCTION

**F**OOD waste is a pressing global issue, with approximately 17% of the world's food production going to waste [1]. In the United States alone, an estimated 30–40% of food goes uneaten annually, resulting in about 160 billion pounds of wasted food [2]. Food waste results in an unnecessary 8–10% increase in greenhouse gas emissions [1] and the loss of land, water, energy, and labor used for farming, transporting, storing, and disposing of food.

A significant fraction of food waste in fresh produce, i.e., fruits and vegetables, occurs at the retail and consumer levels, up to 31% according to the USDA Economic Research Service [3]. This is partly due to the lack of cost-effective and scalable solutions that retailers and consumers can use to predict the ripeness

level and remaining lifetime of fresh produce. Specifically, most retailers and consumers still use visual (e.g., color) and/or tactile (e.g., firmness) inspection to assess the ripeness level of fresh produce, which is a slow process with limited accuracy [4]. This means retailers may discover very late that a batch of fruits is near expiration, which leaves little time to offer discounts to accelerate the sale of such fruits, leading to significant food waste and revenue loss. Similarly, consumers may purchase fruits that are not suitable for their use (either close to expiration or not yet sufficiently ripe), likely discarding them.

Multiple biochemical and technological solutions have been proposed to non-invasively estimate the pre- and post-harvest ripeness level of fruits. For example, some fruits emit various amounts of ethylene gas in different stages of their ripeness [5]. Commercial devices, e.g., [6], measure ethylene emission rates, which are then correlated to fruit ripeness. Near-infrared (NIR) spectroscopy has also been proposed for assessing the ripeness of multiple fruits [7]. And, recently, electromagnetic waves in the sub-tera Hertz range (50–600 GHz) have been proposed to estimate the ripeness level of fruits [8], [9]. While these methods offer higher accuracy than manual inspection, they require specialized hardware setups and are too complex and expensive to be used by end consumers and retailers; they are more suitable for food inspection facilities, sorting lines, large warehouses, and processing plants.

In this paper, we address the problem of estimating the ripeness level and remaining lifetime of fruits *using only smartphones*. This is a challenging research problem for multiple reasons. First, the external features and colors of many fruits, e.g., avocados, do not significantly change with ripening. Rather, the changes accompanying ripening, e.g., conversion of starch to sugar, occur *inside* the fruits. Thus, we need to examine the internal changes of fruits *without* damaging them. In addition, chemical changes due to ripening happen gradually, making it hard to use such changes in predicting the remaining lifetime of fruits. Second, fruits exhibit diverse ripening patterns and lifetimes, and a general solution should account for these differences. Third, the characteristics of smartphone cameras and sensors vary across manufacturers and models. Also, smartphones are used in everyday environments, e.g., grocery stores and homes, which unlike laboratories, have uncontrolled and diverse illumination. The wide diversity of smartphones and illuminations further complicates tracking changes inside fruits.

We propose a cost-effective solution for assessing fruit ripeness and remaining lifetime using smartphones. Our solution analyzes the optical properties of fruits in different ripening

Received 20 September 2024; revised 19 July 2025; accepted 30 July 2025. Date of publication 19 August 2025; date of current version 3 December 2025. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Recommended for acceptance by Z. Li. (*Corresponding author: Mohamed Hefeeda.*)

The authors are with the School of Computing Science, Simon Fraser University, Burnaby, BC V5A 1S6, Canada (e-mail: mhefeeda@sfu.ca).

This article has supplementary downloadable material available at <https://doi.org/10.1109/TMC.2025.3599917>, provided by the authors.

Digital Object Identifier 10.1109/TMC.2025.3599917

stages without damaging them. It then maps these properties to easy-to-understand categories by consumers and retailers, such as Unripe, Ripe, and Expired. Our solution also estimates the remaining lifetime of fruits, enabling retailers and consumers to minimize food waste. Our rigorous experimental study demonstrates the accuracy of the proposed solution.

Specifically, the contributions of this paper are:

- We conduct spectral analysis of different fruits throughout their lifetime using a hyperspectral camera in Section IV. Our analysis shows the limitations of relying only on external visual features to assess fruit ripeness, and it demonstrates the feasibility of tracking chemical changes occurring inside fruits using signals in the 400–1000 nm range, which is the same range of camera sensors on smartphones.
- We present a method for conducting spectral analysis to assess fruit ripeness and remaining lifetime on smartphones in Section V.
- We analyze the auto-catalytic ethylene production process that accompanies fruit ripening, and we define intuitive ripeness and lifetime labels based on this analysis in Section VI.
- We implement a mobile application, RipeTrack, to demonstrate the practicality of the proposed approach in Section VII.
- We conduct an extensive evaluation study to analyze the accuracy, robustness, and extensibility of RipeTrack in Section VIII. Our results show, for example, that RipeTrack can identify the ripeness level of avocados and pears with an accuracy of 95% and 98%, respectively, and it can predict their remaining lifetimes with an accuracy of 93% and 95%. Our results also demonstrate the robustness of RipeTrack to diverse illuminations and smartphones. Further, we show the generality of RipeTrack by extending its functionality to new fruits using transfer learning, including e.g., bananas, mangos, and nectarines, and we test it in multiple grocery stores.

To the best of our knowledge, this is the first work that assesses fruit ripeness using *only* smartphones. Recent works [8], [9] utilize sub-tera Hertz waves, which are not available on smartphones. Other works for food analysis, e.g., [10], [11], [12], [13], do not address fruit ripeness. For example, LiqRay [10] and RF-EATS [11] identify liquids, LiquidHash [12] detects adulteration in liquids, and CapCam [13] tests water contamination. We summarize the related works in Section II. The code and datasets of this work are **open source** [14].

## II. BACKGROUND AND RELATED WORK

*Fruit Types and Ripening Process:* There are two broad categories of fruits: climacteric and non-climacteric. Climacteric fruits, such as pears, apples, avocados, and bananas, continue their ripening process after they are harvested from their plants. Non-climacteric fruits, such as grapes, strawberries, and blueberries, stop ripening once harvested. We focus on climacteric

fruits, as non-climacteric fruits are typically ripe by the time they reach grocery stores.

Assessing fruit ripeness is a complex problem, as it depends on many factors, including external features such as shape, firmness, and color, as well as internal features such as moisture content, soluble solids, acidity, and sweetness. At a high level, as the fruit ripens, it transforms from being hard, sour, often greenish, and odorless to soft, sweet, colorful, and fragrant. These changes are due to the various chemical reactions that occur mainly *inside* the fruit. For example, starch molecules are converted into sugars during ripening. Cell walls of the fruit begin to degrade, which makes it softer and changes its moisture content.

*Ripeness Metrics:* Multiple objective metrics have been developed to measure various aspects of fruit quality and ripeness [7], [15], [16], including Dry Matter (DM), Titratable Acidity (TA), Oil Content, and Total Soluble Solids (TSS or Brix). These metrics help in crucial aspects of fruit farming and handling, such as determining the ideal time to harvest, sorting fruits based on the characteristics needed for certain products (e.g., sweetness level), and adjusting storage environmental conditions to suit different fruits. However, devices that measure these metrics, e.g., [17], are typically expensive, complex to set up and operate, and/or require elaborate calibrations. In addition, these metrics are less useful for end consumers and retailers, who are mostly interested in more direct metrics, such as the remaining lifetime of fruit, and intuitive classification, such as Unripe/Ripe/Expired. We define and measure such metrics.

*Assessing Fruit Ripeness:* Fruit ripeness can be assessed at two main stages: (i) pre-harvest to decide the ideal time to harvest and (ii) post-harvest to track the suitability of fruits for consumption. The post-harvest stage has multiple sub-stages, including transportation, storage, display at retailers, and use by consumers. We focus on tracking fruit ripeness for retailers and consumers in the post-harvest stage, where up to 31% of fruits are wasted [3].

Traditional approaches, which are still in use by many retailers and consumers, manually assess fruit ripeness based on features such as color and firmness [4]. For example, the guidelines in [15] provide Color Gauges for evaluating the quality and ripeness of fruits such as tomatoes and apples. Comparing fruit colors against such charts is neither easy nor accurate, especially under different lighting conditions in homes and grocery stores. Rizzo et al. [16] summarize automated approaches that utilize machine learning to assess ripeness using images captured by regular RGB cameras. However, as demonstrated in Section IV, the external colors and visual features may not accurately reflect the chemical changes happening inside some fruits.

To enable tracking internal changes in fruits, several works have proposed using NIR signals that can penetrate fruit surfaces. For example, Olarewaju et al. [18] use a bench-top spectrometer operating in the 700–2500 nm range to measure dry matter and oil content in avocado to predict its ripeness. The survey in [7] summarizes recent methods that utilize NIR signals to measure various metrics, e.g., DM, Brix, and TA. These methods, however, are tightly coupled with the considered

fruit and metric(s), and thus, they are hard to generalize to other fruits or even to other varieties of the same fruit.

Finally, AgriTera [8] and Meta-Sticker [9] propose using sub-tera Hertz waves to estimate fruit ripeness in terms of DM and Brix. However, sub-tera Hertz signals are not currently available on smartphones.

*Other Food Analysis Systems:* Multiple works have proposed systems to analyze various aspects of foods [10], [11], [12], [13], [19], [20], [21], [22]. LiquidHash [12] models the motion of air bubbles inside bottles to detect adulteration in liquids. CapCam [13] estimates the surface tension of liquids to identify alcohol concentration and water contamination levels. Vi-Liquid [19] identifies liquids by measuring their viscosity coefficients using phone accelerometers. RF-EATS [11] and LiqRay [10] utilize RFID tags to distinguish between various liquids. PowDew [20] detects counterfeit infant formula by analyzing the interaction of water droplets with the powdered formula using smartphones. MeatSpec [21] uses multi-spectral cameras to detect meat adulteration. MobiSpectral [22] identifies organic fruits using spectral analysis. Similar to the work in this paper, MobiSpectral reconstructs the spectrum using a machine learning model, which is based on [23]. We compare the proposed reconstruction model against the one in MobiSpectral and show that our model is more efficient. For example, in inference mode, it runs 30 times faster and requires 2.4 times less memory compared to MobiSpectral, which is crucial for mobile platforms with limited resources. Our model also produces better reconstruction results, as shown in Section VIII-C.

*Summary:* Prior works for assessing fruit ripeness utilize expensive devices that require special setups and calibrations, which make them more suitable for inspection laboratories and large manufacturing facilities. Further, most of these works are designed for a specific fruit or small group of similar fruits. In contrast, RipeTrack is designed for consumers and retailers, uses only smartphones, works in diverse and practical environments, provides intuitive ripeness metrics, and can easily be extended to different fruits.

### III. PROBLEM DEFINITION AND CHALLENGES

The problem addressed in this paper is how to determine the ripeness level and remaining lifetime of climacteric fruits using smartphones operating in regular environments such as grocery stores and homes without damaging these fruits. We summarize the challenges of tackling this problem in the following.

*Non-destructively Tracking Internal Changes:* As mentioned above, fruits undergo chemical changes during the ripening process, which transform some materials into others, e.g., starch to sugar, and alter the water and solid contents of the fruits. To track these internal changes without damaging the inspected fruits, we propose using spectral analysis, which can identify materials based on their electromagnetic properties [24]. This, however, is a challenging task because the internal fruit changes occur gradually over a period of time, and more importantly, the output organic materials from these changes are not substantially different from the input materials in terms of spectral analysis. That is, the differences in the spectral characteristics

of organic materials found in fruits are very subtle. Thus, in Section IV, we first conduct an experimental study to investigate the feasibility of assessing fruit ripeness using spectral analysis utilizing a hyperspectral camera that captures more than 200 bands. Our analysis shows that spectral signatures can be created to represent the chemical compositions of fruits throughout their lifetimes, which provides accurate information for assessing their ripeness.

Hyperspectral cameras are expensive (tens of thousands of dollars) and require strict illumination (halogen sources), which is hard to achieve in environments such as grocery stores and homes. In Section V, we propose a method to conduct spectral analysis on phones. This method uses various signals captured by phones and *upscales* them into spectral bands similar to the ones captured by hyperspectral cameras, which provides the needed information to assess fruit ripeness.

*Difficulty of Determining Ripeness Level:* The ripeness level of some fruits, e.g., bananas, can be estimated from their color. However, the external appearance of many other fruits, e.g., avocados and green apples, does not significantly change with time, which makes it harder to assess their ripeness level. Further, even for fruits that do change colors, the changes could be difficult for inexperienced consumers to detect. For example, some types of pears gradually change their color from greenish to yellowish as they ripen, which is not easy to distinguish, especially in low lighting. In Section VI, we present our approach for modeling the ripeness level and fruit lifetime based on the emission rate of the ethylene gas that accompanies the ripening process.

*Diversity of Phones and Illuminations:* To be of practical value, a mobile application for ripeness analysis must function in everyday environments such as grocery stores and homes. These environments, however, have quite diverse illumination sources, including LED with different color temperatures, fluorescent, sunlight, and arbitrary mixtures of these sources. In contrast, inspection facilities, where spectral analysis is typically performed, have strict illumination conditions. In addition, the application should work with various phones that may have different resolutions and processing steps for RGB images, e.g., white balancing, demosaicing, and color transformation. NIR camera systems on phones may also operate in different wavelengths (between 940 and 980 nm) and resolutions. The diversity in phones and illuminations negatively impacts the accuracy of the spectral analysis, as such analysis relies on detecting small variations of the reflected signals from the scene. In Section V, we present methods to handle this diversity and improve the robustness of the proposed system.

### IV. TRACKING INTERNAL CHANGES IN FRUITS USING HYPERSPECTRAL CAMERAS

Light wavelengths penetrate fruit surfaces at different depths [25]. While this penetration is at the millimeter scale, it provides valuable information for analyzing the fruits. This is the basis of spectral analysis of fruits in general.

Prior works have shown that spectral analysis in various ranges of the spectrum can identify organic materials. For

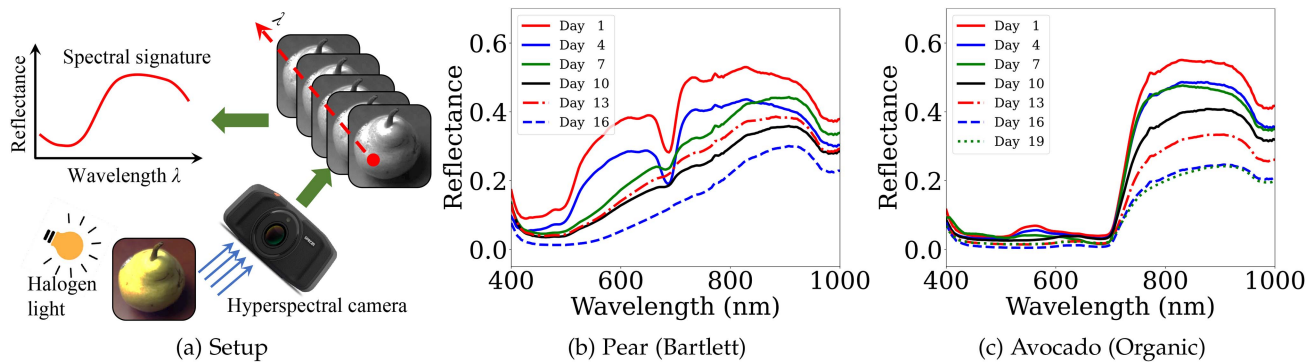


Fig. 1. Spectral analysis of fruits over their lifetime using a hyperspectral camera in the 400–1000 nm range.

example, soluble solids, e.g., sugars, can be observed in the 750–1100 nm range [26], oil content in avocados can be measured in the 2200–2400 nm range [18], water content can be discerned in the 960–980 nm range [27], and the breaking down of Chlorophyll into pigmentation can be tracked in the visible light (400–700 nm) range [15]. Many of the previous works, however, use hyperspectral cameras or spectrometers operating in ranges that extend beyond the available range in smartphone cameras, which is 400–1000 nm. In addition, each of these works focuses on a specific organic material, which may not generalize to other fruits.

In this section, we conduct experiments to demonstrate the feasibility of assessing fruit ripeness and remaining lifetime using a hyperspectral camera that operates in the 400–1000 nm range, which has not been done before in the literature. Our goal is to provide a general framework for analyzing the spectral characteristics of different fruits over their lifetime.

Fig. 1(a) shows our experimental setup. The model of the hyperspectral camera is Specim IQ. The scene is illuminated using a halogen light source, following the recommendations of the camera’s manufacturer. The camera captures 204 spectral wavelengths (aka bands), each with a spatial resolution of  $512 \times 512$  pixels. Thus, the output of this camera is 3-D hyperspectral images with dimensions of  $512 \times 512 \times 204$ , providing spatial details of objects in the captured scene as well as how they reflect different wavelengths in the spectral domain. The normalized reflectance across wavelengths is known as the *spectral signature*, which is computed per pixel.

We analyze the spectral signatures of several fruits throughout their lifetimes, including Pear (Bartlett), Pear (Bosc), Avocado Hass, Avocado (Organic), Mango, and Banana. As detailed in Section VIII-A, we coordinated with local grocery stores to obtain fruit samples on the same day they were delivered, and we cross-checked the observed lifetimes versus the expected ones reported in the food science literature. We kept the samples in our lab, which has temperature, light, and humidity levels similar to those found in homes and grocery stores where such fruits are typically displayed and stored. For each fruit sample, we captured a hyperspectral image every 24 hours using the same conditions (i.e., halogen light source with the same intensity and camera mounted on a tripod to ensure the same capturing distance and angle). We kept capturing hyperspectral images of

the fruits until they expired. These experiments lasted close to 40 days.

For every fruit, we compute a signature for each day of its lifetime. We present representative signatures of Pear (Bartlett) and Avocado (Organic) in Fig. 1; other fruits exhibit similar patterns. These two fruits differ significantly in terms of shape, color, texture, and lifetime. To avoid cluttering the figures, we plot signatures every three days.

Let us first focus on the spectral signatures of pears in Fig. 1(b). As pears ripen, their exterior color gradually changes because chlorophyll (the greenish color) breaks down into new pigments (yellowish-reddish). These changes can be tracked by the reflectance level in the visible light range between 400 and 700 nm. In addition, during the ripening process, the water content increases due to the chemical reactions that break down starch into simpler sugars. Changes in the water content across different days can be seen in the right part of the figure, around the 960–980 nm range. Furthermore, the increase of water over time leads to more light absorption by the fruit across most wavelengths, which is shown in the figure by the more flattened curves with lower reflections in the later days of the fruit’s lifetime. Compare, for instance, the signatures of Day 1 and Day 7 around the 690–710 nm range and the disappearance of the curve dip over time around that range. This occurs because the new color reflects more light in that range.

Unlike pears, avocados do not significantly change their external color as they ripen. This is shown in Fig. 1(c), where the reflectance in the visible range is almost constant and close to zero, as avocados have a dark color that absorbs most visible wavelengths. Thus, the visible range of the spectral signatures provides limited help in assessing the remaining lifetime and ripeness level of avocados. However, the NIR range, between 700 and 1000 nm, reveals noticeable differences between the spectral signatures of avocados across days. For example, the water content increases with time in avocados, leading to more flattened curves with lower reflections in the later days of the fruit’s lifetime.

**Summary and Proposed Framework:** The above experiments reveal subtle differences among spectral signatures of the same fruit computed at different points in its lifetime. Some of these differences can be seen in the visible light range, while others can only be detected in the NIR range. All experiments were

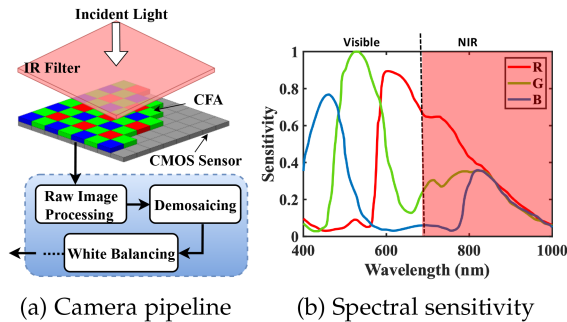


Fig. 2. Simplified illustration of smartphone cameras and their spectral sensitivity. The IR filter removes all signals beyond 700 nm.

conducted using a hyperspectral camera operating in the 400–1000 nm range.

Thus, instead of analyzing separate organic materials, e.g., sugar, water, and oil, of individual fruits as done in prior works [15], [18], [26], [27], we propose constructing spectral signatures that represent the whole structure and chemical composition of fruits at different points in their lifetimes. The shape of the spectral signatures and how they change over time can provide rich enough information for assessing the ripeness and remaining lifetime of different fruits, alleviating the need to precisely measure the presence/concentration of various organic materials in the fruits, which may require complex devices and does not generalize to different fruits.

## V. TRACKING INTERNAL CHANGES IN FRUITS USING SMARTPHONES

### A. Overview and Limitations of Smartphones

The spectral analysis in Section IV was conducted with an expensive (30,000 USD) hyperspectral camera and under ideal (halogen) lighting conditions. Such cameras have complex hardware, e.g., collimating lenses and light dispersion components, to capture the scene across 200+ bands. Our goal is to produce comparable spectral signatures utilizing phones working in arbitrary lighting conditions and then use these signatures to analyze the remaining lifetime and ripeness level of fruits. This, however, is a complex problem for multiple reasons.

The first reason is that smartphone cameras are much simpler than hyperspectral cameras. As illustrated in Fig. 2(a), a smartphone camera utilizes a color filter array (CFA) that enables the 2-dimensional CMOS sensor to capture light in the Red (R), Green (G), and Blue (B) wavelength bands. The most common CFA is the Bayer pattern shown in the figure, where more pixels are allocated to the green band than to the blue and red bands, since the human visual system is more sensitive to green color. Each pixel (photodiode) on the CMOS sensor captures only one of the three colors according to the filter pattern and converts it into an electrical signal. The electrical signals from all pixels are then processed (e.g., amplified and digitized) by the Raw Image Processing module. Then, a Demosaicing algorithm is used to interpolate the other two colors based on the neighboring pixels. Then, additional steps, such as white balancing, color correction,

and compression, are performed to produce the output RGB image. The simple design of smartphone cameras allows them to capture only three bands, unlike hyperspectral cameras that capture more than 200 bands.

The second reason is that smartphone cameras use infrared (IR) filters to remove all signals in the 700–1000 nm range to avoid over-saturating the red band and damaging the visual quality of RGB images. As discussed in Section IV, the NIR range is essential for fruit ripening analysis because signals in this range can penetrate the fruit surface and reveal changes happening inside the fruit. The third reason is the diversity of smartphone cameras and the arbitrary illumination of the environments in which the smartphones operate, as we discussed in Section III.

*In summary*, due to their relatively simple design, smartphone cameras produce coarse-grained information, which is insufficient for spectral analysis, as it requires many bands to create spectral signatures.

To address these challenges and enable conducting spectral analysis on smartphones, we propose *upsampling* the few captured bands by smartphone cameras to many bands. This is sometimes referred to as spectral reconstruction in the literature [23], [28], [29]. In the **Supplementary Materials** (Section A.1), we analyze the state-of-the-art reconstruction model [23] for the suitability of conducting spectral analysis of fruits. Our analysis shows that this model produces low-quality reconstructed bands in the NIR range, leading to significant errors that compromise the accuracy of spectral signatures derived from these bands. One of the main reasons behind this poor performance is the absence of any NIR signals in the input, which causes the model to hallucinate bands in the NIR range. To address this problem, we present three possible solutions to obtain NIR signals in Section V-B. Then, in Section V-C, we present a reconstruction model that produces more accurate bands in the 400–1000 nm range and is robust to practical illuminations, which allows the model to be used for mobile applications in everyday environments. In the evaluation (Section VIII-C), we show that the proposed model outperforms the closest model in the literature [22].

### B. Acquiring NIR Signals on Smartphones

We present three practical solutions to obtain NIR signals on smartphones. The first one uses the NIR camera on modern smartphones, similar to [22]. Specifically, many recent smartphones, e.g., Google Pixel 4, Apple iPhone X, Samsung Galaxy S8, Huawei Mate 20, and their sequels, contain NIR cameras. NIR cameras are usually used for face identification and depth estimation. Depending on the manufacturer, the NIR camera uses a single band in the 940–980 nm range. Smartphones with NIR cameras come with illumination sources in the NIR range. These sources project invisible waves on objects, which are reflected and captured by the NIR camera. This solution, which we refer to as RGB+NIR, does not require any hardware changes.

The second solution is to remove the IR filter shown in Fig. 2(a). Alternatively, commercial camera modules that do not

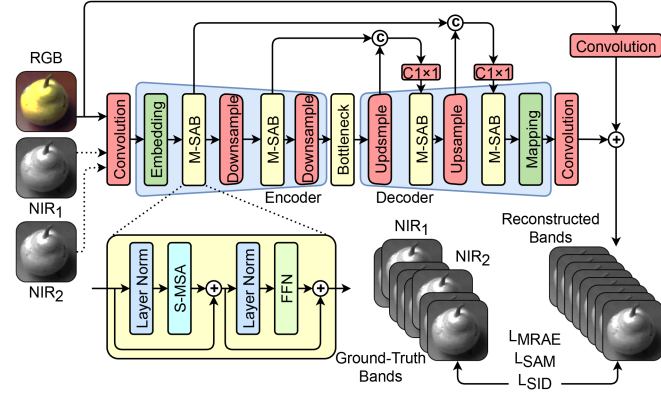


Fig. 3. Architecture of the proposed spectral reconstruction model.  $L_{MRAE}$ ,  $L_{SAM}$ , and  $L_{SID}$  are loss functions to improve the reconstruction accuracy in the spectral and spatial domains. S-MSA is a Spectral-wise Multi-head Self-Attention module, and FFN is a Feed Forward Network.  $NIR_1$  and  $NIR_2$  are bands in the 940–980 nm range used in training to improve robustness.

come with IR filters, such as the Raspberry Pi Camera Module 3 [30], can be used. This solution, which we refer to as No IR Filter, is more suitable for imaging systems designed for specialized tasks, such as quality inspection devices.

The third solution to obtain NIR signals on smartphones is to change the CFA to have an explicit filter for the NIR channel. Designing a custom filter array is a complex research problem in its own right, as there are numerous possible filter patterns, and the performance of each pattern depends on several factors, including CMOS sensor sensitivity and the processing steps performed on the sensor's output. Monno et al. [31] conduct a comprehensive performance analysis of various color filter arrays. They present a  $4 \times 4$  filter design that yields the best overall performance in terms of achieving an explicit NIR band with minimal impact on the RGB bands. This design, however, was analyzed using only *synthetic* data, not real camera sensors. We identified and purchased a commercial camera sensor with a similar filter pattern, which is the CMOS Image Sensor Model AR0237 RGB-IR from ON Semiconductor [32]. This solution, which we refer to as Custom Filter, can be useful for designing future smartphones and specialized imaging systems for quality analysis and inspection.

### C. Spectral Reconstruction Model

The proposed spectral reconstruction model is shown in Fig. 3, which is designed using vision transformers [33] similar to recent works, e.g., [23]. Compared to prior works, however, our model considers the NIR band as an additional input, introduces new loss functions to enhance accuracy, improves robustness to diverse phones and illuminations, and significantly reduces memory requirements and training and inference times—all are critical factors for phones.

Vision transformers can efficiently learn correlations in the input data through a mechanism known as self attention [34]. They divide an image into non-overlapping patches, map these patches to vectors, and encode the positions of patches into vectors as well. Then, vectors representing patches and their positions are passed through an encoding stage, where the self-attention

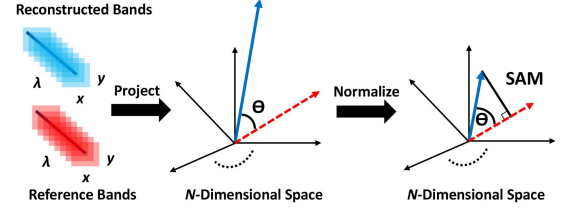


Fig. 4. Illustration of the SAM loss function.

module captures the correlations among patches. Typically, multiple self-attention modules are applied in parallel to capture various patterns and semantic relationships across patches. This attention focuses on capturing the *spatial* relationship among pixels within the image, which is useful for computer vision tasks such as image segmentation and classification. For spectral analysis, however, the *spectral* relationship is also important.

**Improving Reconstruction Accuracy:** To make the reconstruction model consider both the spatial and spectral domains, we present two optimizations. For the first optimization, we adopt the Multi-head Spectral-wise Attention Block (M-SAB) proposed in [23], which computes the attention across spectral bands.

For the second optimization, we propose a loss function with three components: (i) Mean Relative Absolute Error (MRAE), (ii) Spectral Angle Mapper (SAM), and (iii) Spectral Information Divergence (SID). MRAE measures the absolute relative error between pixel values of the reconstructed and ground truth bands. It strives to ensure the accuracy of the reconstructed bands in the spatial domain, and it is computed as:

$$L_{MRAE} = \frac{1}{HWN} \sum_{x=1}^H \sum_{y=1}^W \sum_{\lambda=1}^N \left| \frac{\hat{X}(x, y, \lambda) - X(x, y, \lambda)}{X(x, y, \lambda)} \right|, \quad (1)$$

where  $N$  is the number of bands,  $H$  and  $W$  are the spatial resolution, and  $\hat{X}$  and  $X$  represent the reconstructed and ground-truth bands, respectively.

SAM measures the similarity between two spectra by computing the angle between them [24]. Fig. 4 illustrates the SAM metric, where the reconstructed and ground truth bands are first projected and normalized as vectors in the  $N$ -dimensional space, and then the angle between these vectors is computed. SAM is computed as:

$$L_{SAM} = \frac{1}{HW} \sum_{x=1}^H \sum_{y=1}^W \cos^{-1} \left( \frac{\sum_{\lambda=1}^N \hat{x}_{\lambda} x_{\lambda}}{\sqrt{\sum_{\lambda=1}^N \hat{x}_{\lambda}^2} \sqrt{\sum_{\lambda=1}^N x_{\lambda}^2}} \right), \quad (2)$$

where  $x$  and  $\hat{x}$  are spectral vectors from the reconstructed and ground truth bands.

SID measures the difference between the probability distributions of the reconstructed and ground truth bands [35]. SID first transforms the spectral bands into probability distributions, and it then calculates the difference between them, as illustrated

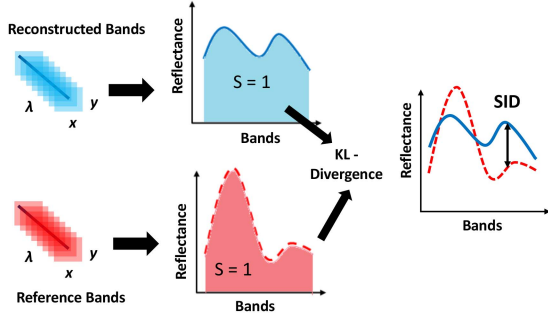


Fig. 5. Illustration of the SID loss function.

in Fig. 5. SID is computed as:

$$L_{SID} = \sum_{\lambda=1}^N p_{\lambda} \log \left( \frac{p_{\lambda}}{q_{\lambda}} \right) + \sum_{\lambda=1}^N q_{\lambda} \log \left( \frac{q_{\lambda}}{p_{\lambda}} \right), \quad (3)$$

where  $p$  and  $q$  are the normalized vectors of the reconstructed and ground truth bands.

SAM and SID strive to make the reconstructed bands as close as possible to the ground-truth bands across the spectral domain.

The total loss function in our model is given by:

$$L = L_{MRAE} + w_1 \times L_{SAM} + w_2 \times L_{SID}, \quad (4)$$

where  $w_1 = 0.1$  and  $w_2 = 0.001$ , which are selected to ensure SAM and SID do not overpower the other losses as they could induce distortions in the reconstructed bands if their weights are high [36].

*Improving Robustness:* Hyperspectral applications track the value of each pixel across different bands to create spectral signatures. Pixel values, however, depend on the illumination of the scene and the camera hardware. This means various cameras and illuminations may result in different spectral signatures of the same scene, limiting the practicality of the proposed approach. To address this critical issue, we divide the problem into two parts corresponding to the model's inputs: NIR and RGB images.

NIR cameras on recent phones may not use the same wavelength. Instead, they pick an operating point in the 940–980 nm range, leading to differences in NIR images captured by different phones. To mitigate this problem, we train the model to reconstruct spectral bands with different NIR images as input. This is illustrated in Fig. 3, where we pair an input RGB image with multiple NIR images at different wavelengths instead of only one. We call this approach NIR data augmentation. That is, a single pair of (RGB, NIR) images is expanded to  $L$  pairs of (RGB, NIR<sub>1</sub>), (RGB, NIR<sub>2</sub>), ..., (RGB, NIR<sub>L</sub>) images, where the NIR images have different wavelengths. Then, the model is trained to produce the same reconstruction results for all image pairs.

Unlike NIR images, RGB images are affected by the scene's illumination, in addition to their dependence on the camera hardware. Specifically, most phone manufacturers implement various *proprietary* algorithms in the processing pipeline to enhance the visual appearance of the final images. This means the processing pipeline varies across cameras. In addition, some

essential steps, e.g., white balancing, estimate the illumination of the scene and adjust the colors of images accordingly. The variability of RGB images produced by different phones and under various illuminations significantly reduces the accuracy of the reconstructed bands.

To address this problem, we employ an image normalization approach similar to [22], where we implement a deep-learning model that maps an input RGB image to a common representation, regardless of the camera's characteristics and scene illumination. All RGB images are first normalized before being used in the reconstruction model. Specifically, the considered image normalization extends the white balancing model in [37]. It transforms all images to a common illumination setting, regardless of the cameras that captured these images and the illuminations used. As in [22], we transform all images to the daylight illumination (5500 Kelvin) setting.

## VI. DETERMINING GROUND-TRUTH RIPENESS

The goal of this section is to develop an accurate and intuitive method for labeling the ripeness levels and remaining lifetime of fruits. We base our method on an established body of research in food science. In particular, the emission rate of ethylene has been established for decades as a robust indicator for fruit ripening [5]. Briefly, a small amount of ethylene is generated when a fruit starts to ripen after harvesting. Then, ethylene catalyzes multiple chemical reactions in the fruit, which helps the ripening process. These reactions, in turn, produce more ethylene, which further catalyzes more chemical reactions and accelerates ripening. This is known as the auto-catalytic production of ethylene in fruits [38].

To develop our labeling method, we conduct experiments to analyze the ethylene emission rate for different fruits. Our experimental setup is shown in Fig. 6(a), which is similar to setups used in prior works in this domain. The model of the ethylene measurement device is the Forensics Detector FD-90A-C2H4 [39], and it provides an accuracy of 1 ppm (part per million) with a range of 0–100 ppm. The device comes with a probe and gas sampling pump. We place a fruit sample inside a plastic container with a tight lid that has a small opening for the probe. The probe is kept inside the container for 30 seconds, which is the response time of the device. After recording the ethylene emission rate, the fruit sample is removed, and the container is kept open for 3–5 minutes to remove any ethylene traces. Then, the measurement is conducted for another sample of the same fruit. Subsequently, the experiment is repeated for samples of other fruits. Finally, the whole set of experiments is repeated every day around the same time until the fruits expire.

Samples of our results are shown in Fig. 6(b) and (c), where we plot the average ethylene emission rate for every day of the lifetime of pears and avocados. We also plot the confidence interval for each day as error bars (average plus/minus one standard deviation). Although pears and avocados have very different lifetimes, their emission curves have the same *pattern* of increasing, then decreasing, and finally stabilizing. Similar emission patterns were observed for other fruits in our study.

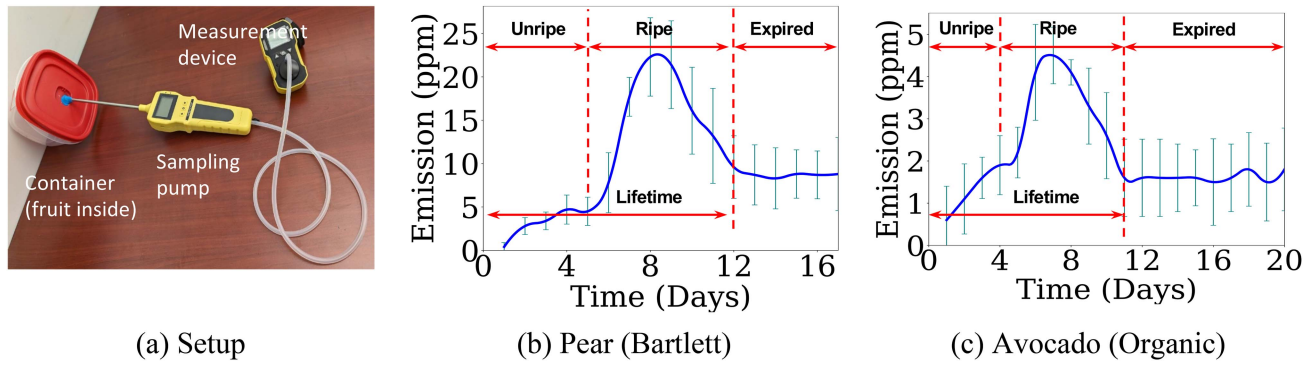


Fig. 6. Analysis of the ethylene emission rate (in parts per million or ppm) for two fruits over their lifetime.

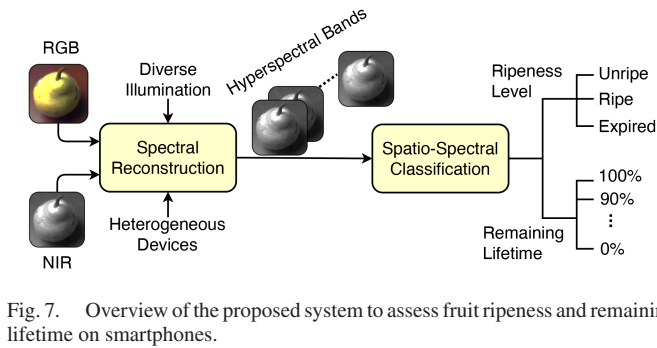


Fig. 7. Overview of the proposed system to assess fruit ripeness and remaining lifetime on smartphones.

Prior food science research, e.g., [40], shows that the unripe stage is characterized by low ethylene emission rates, where ethylene is produced in the so-called ‘auto-inhibitory’ manner. This inhibition continues until the ripe stage starts, where ethylene emission increases rapidly until it reaches its peak in what is referred to as the ‘auto-inductive’ process of ethylene production. After reaching the peak, the ethylene emission rate decreases rapidly until it levels off, indicating the start of the expired stage. Based on this, we define our ground-truth labeling for ripeness level and remaining lifetime. We annotate the curves in Fig. 6(b) and (c) to show the Unripe, Ripe, and Expired stages. We define the *remaining lifetime* as the number of days left until a fruit reaches the beginning of its Expired stage. In our evaluations, we use this ground-truth labeling in training our classification model in Section VII-B. During inference, which is done on smartphones, ethylene measurement is *not* performed; we only use images.

*Alternatives:* The ethylene emission rate provides accurate ripeness and lifetime labeling, but it requires a measurement device. Alternative methods, such as testing fruit firmness and/or matching its color against pre-defined color charts [15], [16], can be used to provide approximate labeling.

## VII. END-TO-END SYSTEM AND MOBILE APP

### A. System Overview and Operation

Fig. 7 provides a high-level overview of the proposed system to assess fruit ripeness and remaining lifetime on phones. It has two deep-learning models for spectral reconstruction and spatio-spectral classification. The system takes as input RGB

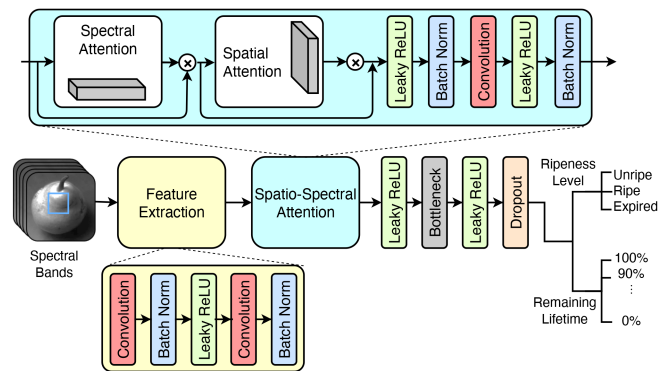


Fig. 8. Design of the proposed spatio-spectral classifier for fruit ripeness level and remaining lifetime.

and NIR images of a fruit captured by phones under arbitrary illumination available in regular environments such as grocery stores and homes. The RGB and NIR images are normalized and fed to the reconstruction model, which produces a configurable number of bands equally spaced in the 400–1000 nm range. In our experiments, we set the number of bands to 68. Reconstructing more bands did not improve the accuracy, while it substantially increased the processing and memory requirements both at the training and inference stages of the two models. The reconstructed bands are given to the classifier, which produces two outputs: ripeness level (Unripe, Ripe, or Expired) and the remaining lifetime as a percentage.

### B. Design of the Spatio-Spectral Classifier

We propose a classification model that considers both the spatial and spectral characteristics of the input bands. This is crucial for identifying the subtle differences among a fruit’s spectral signatures at different points in its lifetime. The proposed model is illustrated in Fig. 8. The input to the classifier is the bands created by the reconstruction model. These bands are fed to a Feature Extraction module to compute low-level features. This is achieved by two convolution layers, interspersed by batch normalization and Leaky ReLU layers. Then, the extracted features are passed to a Spatio-Spectral Attention module, which consists of two attention blocks. The first captures the context

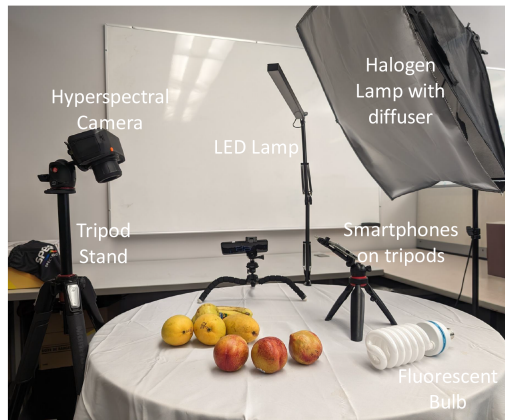


Fig. 9. The testbed used in our experiments.

among the spatial features of bands, whereas the second attends to the spectral features across them.

The classifier produces outputs in two categories: ripeness level and remaining lifetime. The ripeness level can be Unripe, Ripe, or Expired. The remaining lifetime is represented as a percentage instead of an absolute value, which enables generalization to various fruits with diverse lifetimes. We configure the classification model to produce 11 classes for the remaining lifetime: 0%, 10%, ..., 100%. We believe this is a sufficient granularity, as the lifetime of most fruits ranges between one to three weeks. Thus, these 11 classes allow predicting the remaining lifetime in a granularity of 1–2 days. Nonetheless, the model can easily be configured to produce different numbers of classes.

### C. Mobile App

We have developed an Android application as a proof of concept; sample screenshots are given in Fig. 20, and the code can be found at [14]. The application is written in Kotlin and compiled using Gradle version 8.0. The image capturing modules of the application are built on top of the Android Camera2Basic Project.

The spectral reconstruction and classification models are developed in PyTorch. They are first trained on a workstation. Then, the trained models are quantized using the default quantization parameters in PyTorch, which represent 32-bit floating-point weights as 8-bit integers. This makes the models run 2–4X faster while having a small impact on the accuracy. While quantization of the models is optional, we opted to utilize it to enable RipeTrack to function on many phones with limited computing resources. Then, the quantized models are ported to the Android platform using the PyTorch’s JIT Trace. Finally, the trained and quantized models are integrated with the mobile application for inference.

The application can access both the phone’s RGB and NIR cameras through Android’s Camera API2. It captures the NIR image immediately after capturing the RGB image of the scene. The RGB and NIR images are then passed to the reconstruction model, which creates 68 spectral bands. The reconstructed bands are then passed to the classification model, which produces the

ripeness level and remaining lifetime. The application stores intermediary data, e.g., reconstructed bands, for debugging and further analysis.

### D. Limitations and Extensions

RipeTrack requires NIR signals, and we presented three solutions to acquire such signals. One of these solutions offers good accuracy without requiring modifications to phones. It requires accessing the NIR camera, which is available in many recent phones. However, some manufacturers, e.g., Apple, do not currently allow external developers to access the NIR camera. The models of RipeTrack need to be trained. Our open-source hyperspectral imaging dataset [14] provides a starting point. To support new fruits, a few hyperspectral images would need to be captured and used to fine-tune the models. In addition, ground-truth labels for ripeness and lifetime need to be defined by measuring ethylene emission; alternatively, they can be approximated using manual methods, such as comparing fruit colors with standard charts and/or performing firmness tests. Finally, RipeTrack is designed for climacteric fruits. It is not suitable for non-climacteric fruits such as grapes, strawberries, and blueberries. These types of fruits stop ripening after being harvested, unlike the climacteric ones.

## VIII. EVALUATION

We first describe our setup and datasets in Section VIII-A. Then, we demonstrate the accuracy of the proposed three methods for conducting spectral analysis on phones in Section VIII-B. In Section VIII-C, we compare the reconstruction model of RipeTrack to the state-of-the-art. Then, we show the accuracy of RipeTrack in assessing ripeness levels and remaining lifetime for different fruits and its extensibility to new fruits in Section VIII-D. In Section VIII-E, we analyze the performance impact of various components of RipeTrack and demonstrate its robustness to diverse illuminations, phones, and capturing distances. We also analyze multiple system parameters, e.g., training and inference times, and we test RipeTrack in five different grocery stores, demonstrating its practicality.

We share our codes and datasets with the research community at [14], with details to reproduce our results.

### A. Experimental Setup

*Testbed:* Fig. 9 shows our testbed, which consists of:

- *Hyperspectral Camera:* Used to capture hyperspectral images of fruits for evaluating the spectral reconstruction model. The camera model is Specim IQ, which uses a CMOS sensor operating in the 400–1000 nm range. It captures 204 bands with a spectral resolution of 3 nm. The spatial resolution of each band is  $512 \times 512$  pixels. Thus, the output of this camera is images with dimensions of  $512 \times 512 \times 204$ . The camera takes about 180 seconds to capture a single image because it linearly scans the scene.
- *Camera Module with Custom RGB+NIR CFA:* Used to evaluate the spectral reconstruction model. The model is ON Semiconductor AR0237 [32].

TABLE I  
SUMMARY OF THE FRUIT RIPENING DATASET

|                 | Fruit Samples | Hypersp. Images | Ethylene Readings | Observed Lifetime |
|-----------------|---------------|-----------------|-------------------|-------------------|
| Avocado Organic | 10            | 460             | 230               | 11                |
| Avocado Hass    | 3             | 234             | 117               | 19                |
| Pear Bartlett   | 11            | 382             | 209               | 12                |
| Pear Bosc       | 3             | 276             | 138               | 40                |
| Banana          | 12            | 279             | 168               | 7                 |
| Nectarine       | 3             | 138             | 138               | 16                |
| Mango           | 6             | 144             | 144               | 16                |
| Total           | 48            | 1913            | 1144              |                   |

- *Smartphone without IR Filter:* Used to evaluate the spectral reconstruction model. The model is Google Nexus 5X.
- *Two Unmodified Smartphones:* Used to run RipeTrack and evaluate its accuracy and robustness across different phones. The models are Google Pixel 4XL and OnePlus 8 Pro. Both have RGB and NIR cameras.
- *Various Light Sources:* Used to illuminate the captured scene and evaluate the robustness of RipeTrack under diverse illuminations. The testbed has halogen, LED, and CFL sources.
- *Ethylene Measurement Device:* Used to measure the ethylene emission rate, which defines the ground-truth labeling of the fruit ripeness level and remaining lifetime. The device model is Forensics Detector FD-90A-C2H4 [39], and it provides an accuracy of 1 ppm (part per million) and has a range of 0–100 ppm.

*Fruits Considered:* As summarized in Table I, the considered fruits have diverse external features, colors, ripening patterns, and lifetimes to demonstrate the practicality and robustness of RipeTrack. For example, pear Bosc takes about 40 days to expire, while pear Bartlett expires in about 12 days. Both pears have different colors, and their colors change over time. While the organic and non-organic avocados are hard to distinguish visually, the organic version expires in about 11 days and the non-organic in 19 days. The choice of two varieties of pears and avocados stresses our system, because it would need to learn the internal characteristics of visually similar fruits to predict their ripeness and lifetime. The chosen fruits are among the top items contributing to food waste [3].

*Fruit Ripening Dataset:* We purchased samples of each considered fruit from multiple grocery stores at different times. We coordinated with the stores to obtain these fruits early in their ripening process, typically on the day of their delivery. Although we cannot know exactly when the fruits were harvested, we cross-checked the observed lifetime of each fruit against its expected lifetime in the literature.

This dataset contains hyperspectral images spanning the entire lifetime of fruits and the associated ethylene emission rates. Specifically, we capture two hyperspectral images of each fruit sample every 24 hours, taking them from slightly different angles. We use a halogen light source, as recommended by the camera’s manufacturer. We mount the camera on a tripod and fix the capturing distance throughout the experiments. After

taking the hyperspectral image of a fruit sample, we measure the ethylene emission rate of that sample using the setup illustrated in Fig. 6. We keep capturing images and measuring ethylene until the fruit expires. We associate the ripeness level and remaining lifetime with the captured images at different times using the method described in Section VI.

The data collection process lasted more than *two months* because capturing a single hyperspectral image takes about 3 minutes, and measuring ethylene requires several minutes for each sample. The final dataset has 1,913 hyperspectral images and 1,144 ethylene measurements over the lifetime of seven different fruits, as summarized in Table I. This is a sizable dataset in this domain. Recall that every hyperspectral image has 204 bands, each is a gray-scale image. That is, this dataset has more than 390 K individual images. This dataset is used to train the spectral reconstruction model.

*Mobile Images Dataset:* To realistically evaluate RipeTrack, we collected a dataset using two different phones: Google Pixel 4XL and OnePlus 8 Pro. Google Pixel has resolutions of  $800 \times 600$  and  $640 \times 480$  pixels for the RGB and NIR cameras, respectively, whereas OnePlus has resolutions of  $4032 \times 3024$  and  $2592 \times 1944$  pixels. We scale all RGB and NIR images to  $640 \times 480$  pixels. We captured this dataset while capturing the hyperspectral images dataset for all fruits. Specifically, for each fruit sample, we capture RGB and NIR images using one of the phones. We use illumination sources deployed in real environments (LED and fluorescent). We also use mixtures of these sources and natural sunlight. In total, we captured 3,865 pairs of RGB-NIR images of seven fruits over their entire lifetimes. Out of these pairs, 2,695 were captured using Google Pixel and 1,170 using OnePlus. This dataset is *not* used in training the reconstruction model. It is used only to test the accuracy of estimating fruit ripeness and remaining lifetime.

## B. Accuracy of Spectral Analysis

The reconstruction model produces bands from which we create signatures representing different points in the fruit’s lifetime. Thus, the accuracy of these bands is critical for conducting spectral analysis. We evaluate the accuracy of the reconstructed bands by comparing them against the ground-truth ones captured by the hyperspectral camera.

*Training:* We train three versions of the reconstruction model based on the given inputs. The first version is called No IR Filter, where the input consists of three RGB bands after the IR filter is removed from the camera. The second version takes as input the four bands produced by the custom color filter array and is referred to as the Custom Filter. The third version represents the case of unmodified phones, which take separate RGB and NIR images as input; we refer to this version as RGB + NIR. The fruit ripening hyperspectral images dataset is used to train and test the reconstruction model. It is divided into three partitions: 70% for training, 15% for validation, and 15% for testing. We use images from the first four fruits in Table I in this section, and we keep the others for later testing the extensibility of our model.

TABLE II  
PERFORMANCE OF THE RECONSTRUCTION MODEL

|               | MARE | RMSE | SAM  | SID  | PSNR | SSIM |
|---------------|------|------|------|------|------|------|
| No IR Filter  | 0.14 | 0.03 | 0.10 | 0.04 | 31.2 | 0.95 |
| Custom Filter | 0.08 | 0.01 | 0.06 | 0.01 | 39.4 | 0.99 |
| RGB + NIR     | 0.12 | 0.08 | 0.08 | 0.01 | 34.0 | 0.97 |

Average metrics across all fruits are presented.

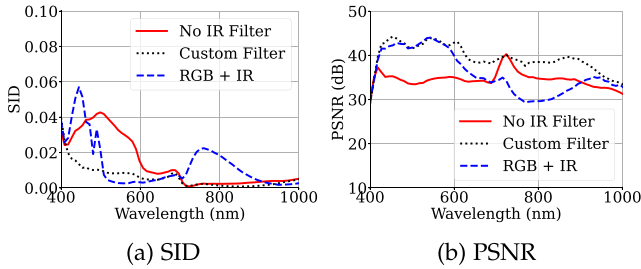


Fig. 10. Band-wise performance analysis of the reconstruction model with different inputs.

We measure the accuracy using six performance metrics commonly used in the literature [22], [24], which are MARE, SAM, SID, RMSE (Root Mean Square Error), PSNR (Peak-Signal to Noise Ratio), and SSIM (Structural Similarity Index Measure). The first four metrics measure the *error* between the reconstructed and ground truth bands from different perspectives. The last two assess the *quality* of the reconstructed bands relative to the ground truth ones. We provide more training details and equations of the performance metrics in Section A.2 in the Supplementary Materials.

**Summary of the Results:** We report the overall performance of the three versions of the reconstruction model in Table II, where we show only averages. Details of individual fruits with confidence intervals are given Section A.3 in the Supplementary Materials. As the table shows, all three versions produce good reconstructed bands. For example, the average PSNR is at least 31 dB, and the average SSIM is close to one. Similarly, the error metrics, especially the crucial SAM and SID error metrics, are close to zero.

To shed light on the relative performance of the three versions, we plot their accuracy across individual bands in Fig. 10 for two representative metrics. The results in Table II and Fig. 10 show that the custom CFA provides the highest accuracy. Surprisingly, using RGB + NIR images, which does not require any phone modification, provides better accuracy than removing the IR filter. This is because, in the first case, the RGB camera still retains the quality of RGB images. This yields a higher reconstruction accuracy in the visible range compared to the No IR Filter case, where the quality of the RGB images is compromised due to interference with IR signals. Further, the additional NIR image, which is around 940 nm, enables the model to reconstruct bands in the 900–1000 nm range with higher accuracy than the No IR Filter case. We note that the RGB + NIR case has less accuracy in the 750–850 nm range, because the transition from the visible range to the NIR range occurs around 700 nm, where all RGB signals are truncated. This provides the reconstruction model

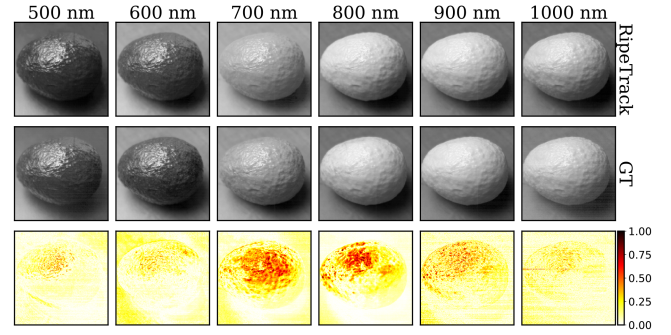


Fig. 11. Visual comparison of the reconstructed bands and the ground truth (GT) ones; the bottom row shows the absolute errors between them as heat maps.

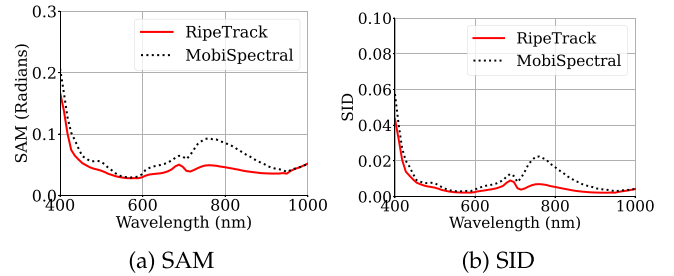


Fig. 12. Performance of the proposed reconstruction model in RipeTrack versus state of the art (MobiSpectral).

with less information to build on in the 750–850 nm range, leading to relatively higher errors.

Finally, we show a few samples demonstrating the quality of the reconstructed bands using RGB and NIR images in Fig. 11. The figure also shows the difference between each reconstructed band and its corresponding ground truth one as a heat map.

### C. Comparison Against State-of-The-Art

The proposed spectral reconstruction model in this paper improves on MobiSpectral [22], by adding multiple loss functions to enhance the quality of the reconstructed bands as well as simplifying the design of the neural network to significantly reduce the computational complexity. MobiSpectral itself was built on the state-of-the-art reconstruction model in [23].

We compare the performance of the proposed reconstruction model against MobiSpectral using the fruit ripening dataset described in Section VIII-A. In both cases, we use RGB and NIR images. We present sample results in Fig. 12 for the two most important metrics: SAM and SID. These results are the averages of the SAM and SID metrics across the test partition of the fruit ripening dataset. The figure shows that the proposed model consistently produces lower SAM and SID (error) values, especially around the 700–900 nm range. This range provides valuable information in the invisible range, which improves the reconstruction accuracy.

In addition, we compare the space and time complexity by running both reconstruction models successively on the same workstation (specs are given in Section VIII-E). The results are summarized in Table III, which shows that the proposed reconstruction model is more efficient than MobiSpectral. For

TABLE III  
COMPUTATIONAL COMPLEXITY OF THE RECONSTRUCTION MODEL IN  
RIPETRACK AND STATE OF THE ART (MOBISPECTRAL)

|                | RipeTrack | MobiSpectral |
|----------------|-----------|--------------|
| Inference Time | 0.11 s    | 3.5 s        |
| GPU Memory     | 3.1 GB    | 10.6 GB      |
| Parameters     | 293,356   | 3,003,708    |

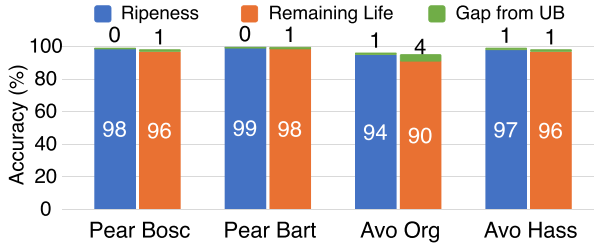


Fig. 13. The accuracy of predicting fruit ripeness and remaining lifetime using images captured by phones.

example, the inference, i.e., reconstructing 68 bands from the input four RGB and NIR images, takes on average 0.11 seconds, which is 30X less than the time needed by MobiSpectral. In addition, the memory footprint of the proposed model is approximately 3.1 GB compared to 10.6 GB for MobiSpectral. These savings in computational resources are crucial when the model is deployed on smartphones with limited resources.

#### D. Accuracy of Assessing Fruit Ripeness

We evaluate the accuracy of determining fruit ripeness and remaining lifetime using the mobile images dataset, which was collected using two unmodified phones under diverse illuminations. This dataset is divided into two partitions: 85% for training and 15% for testing. Data points in the testing partition are from fruit samples that were never seen during training. This stresses our system and shows its robustness to natural variations in samples of the same fruit type. The RGB and NIR images in the training partition are first upsampled to 68 bands using the reconstruction model. The reconstructed bands are then paired with the corresponding ground truth ripeness and remaining lifetime labels to train the classifier.

**Average Accuracy:** In Fig. 13, we summarize the average accuracy of estimating ripeness and remaining lifetime for different fruits. We also measure the accuracy achieved by the expensive hyperspectral camera under ideal (halogen) lighting. In this case, the reconstruction model is not invoked, and the classifier is trained on actual hyperspectral bands. This case represents the *upper bound (UB)* on accuracy achievable through spectral analysis in the 400-1000 nm range. As the results in Fig. 13 show, RipeTrack achieves high accuracy for all considered fruits. Specifically, accuracies of at least 96% and 93% are observed for ripeness and remaining lifetime, respectively. In addition, the accuracy achieved by RipeTrack using phone images is within a few percentage points from the upper bound; percentages are shown on top of the bars. We note that the

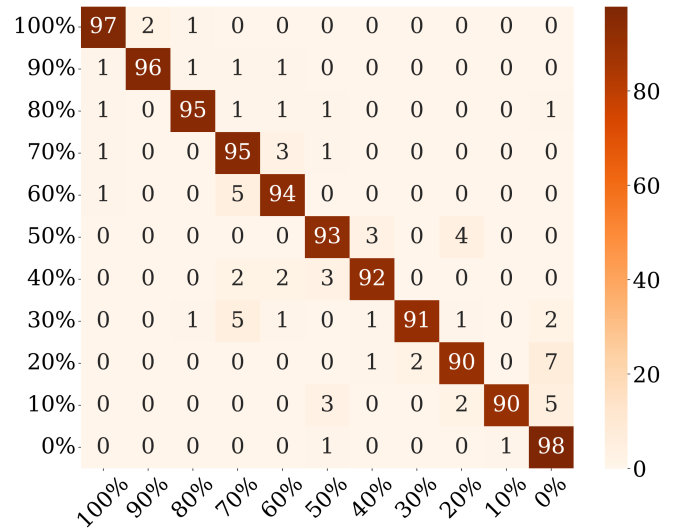


Fig. 14. Accuracy of predicting individual classes of the remaining lifetime. Rows: predicted; Columns: actual.

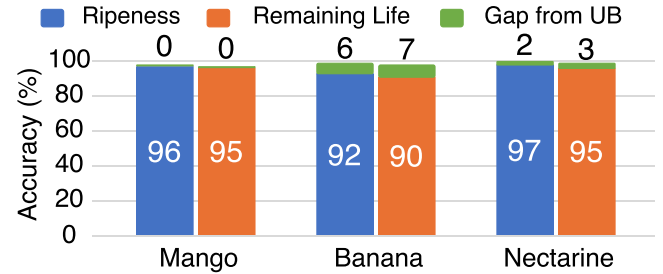


Fig. 15. Extensibility of RipeTrack to other fruits using transfer learning by fine-tuning the model on a few images.

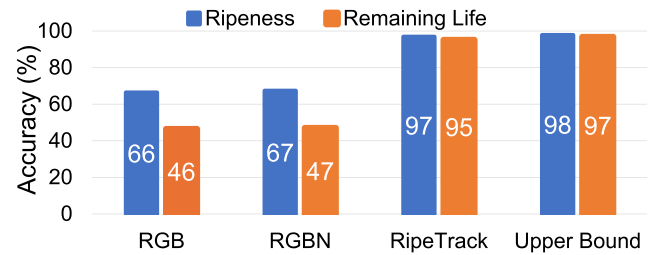


Fig. 16. Ablation study: Performance impact of various components of RipeTrack.

accuracy of assessing avocado ripeness is slightly lower than that of pears, because pears exhibit more external changes during ripening than avocados, which provides additional signals to our models.

**Per-Class Analysis:** We examine the accuracy of predicting individual classes of ripeness and remaining lifetime. A sample of our results is presented in Fig. 14 for the 11 classes of the remaining lifetime; other results are similar. This is a standard confusion matrix computed across all fruits, where each row is normalized and contains the probability distribution of predicting the corresponding label. Values on the diagonal indicate the percentage of predicted labels that equal the true labels. The

figure shows high accuracy across all classes, with classes at both ends of the lifetime achieving relatively higher accuracy, as they are less challenging to identify compared to other classes.

**Extension to New Fruits using Transfer Learning:** The above results were obtained by training and testing RipeTrack on the first four fruits in Table I. We extend RipeTrack to the other three fruits (banana, nectarine, and mango) by fine-tuning its reconstruction and classification models using transfer learning. Specifically, we randomly select a subset (70%) of the hyper-spectral images of these three fruits to fine-tune the reconstruction model. For fine-tuning the classification model, we associate the ground-truth ripeness level and remaining lifetime with the captured images at different times using the method described in Section VI. Similarly, we randomly select 70% of the data of these three fruits to fine-tune the classification model.

We report the average classification accuracy in Fig. 15, which was computed on the remaining (test) data points not used during training. The results confirm the extensibility and accuracy of RipeTrack: It generalized to new fruits with totally different shapes, colors, and chemical compositions than the ones it was trained on by fine-tuning its models on a few fruit samples.

#### E. System Analysis and In-Store Testing

**Ablation Study:** We analyze the performance impact of different components of RipeTrack. Specifically, we evaluate the accuracy of assessing fruit ripeness using only RGB images. That is, the ripeness and lifetime classification model is trained and tested *only* on the RGB images of the mobile images dataset described in Section VIII-A. Then, we add NIR images, but without spectral reconstruction; i.e., we use pairs of RGB and NIR images from the mobile images dataset for training and testing the classification model. Then, we perform spectral reconstruction from RGB and NIR images and use the reconstructed bands to train and test the classification model.

The results of this experiment are shown in Fig. 16, where we also show the upper bound on accuracy obtained by feeding the ground-truth hyperspectral images to the model. As the figure shows, RGB images alone provide low accuracy because they can only model external features of fruits, whereas the ripening process occurs mainly inside the fruits. Using the NIR image helped the performance marginally. This is because the NIR band captured by the phone is fairly narrow and provides limited information to the classification model. A substantial improvement is achieved using the proposed reconstruction model, which brings the accuracy close to the upper bound. The reconstruction model effectively utilizes both the NIR and RGB bands to reconstruct the entire spectrum, providing the model with rich information.

**Robustness to Practical Illuminations:** Our mobile images dataset is captured under diverse illuminations: LED, Fluorescent (CFL), and mixtures of these sources and sunlight (referred to as Mixed). We separate the test partition of the mobile images dataset based on the illumination source. Then, we assess the classification accuracy for each illumination source. The results in Fig. 18 show that the image normalization method of RipeTrack mitigates the differences in illuminations, and the type of illumination does not impact the accuracy. The

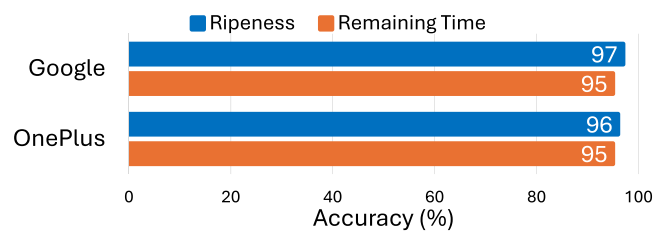


Fig. 17. Robustness and generalizability of RipeTrack across phones. Top row: RipeTrack is trained on images captured by OnePlus and tested on images captured by Google Pixel. Bottom row: the other way around.

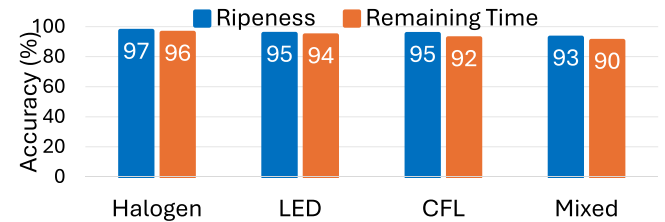


Fig. 18. Robustness to illumination diversity: RipeTrack achieves high accuracy under different illuminations.

accuracy is highest when using halogen sources because they emit power across the entire spectrum, helping the phone capture more reflected signals. Halogen sources, however, are not widely used in homes and grocery stores as they consume substantially more energy than other sources. The mixed scenario is the most challenging, as it includes uncontrolled lighting sources such as sunlight coming from windows and light coming from the normal bulbs in our lab. Nonetheless, RipeTrack still achieves an accuracy of at least 90% in this challenging scenario.

**Robustness and Generalizability of RipeTrack to Diverse Phones:** We separate the images captured by our two phones (Google Pixel and OnePlus) and compute the classification accuracy for each group. Specifically, we train our models on images captured by one phone (e.g., Google Pixel) and test images captured by the other phone (OnePlus). Then, we switch: train on images captured by OnePlus and test on images captured by Google Pixel. These phones have very different camera specifications.

The results are presented in Fig. 17. The first row illustrates the case where the models are trained on images captured by the OnePlus phone and then tested on images captured by the Google Pixel phone. The second row shows the other way around. The results demonstrate the high accuracy achieved by RipeTrack across different phones, showing its robustness and generality. This is achieved by the RGB image normalization and NIR data augmentation methods in RipeTrack, which collectively enable RipeTrack to function on different phones.

**Effect of Capturing Distance:** We analyze the accuracy of RipeTrack when capturing images at different distances. While RGB cameras can capture images multiple meters away, phone NIR cameras typically have much smaller operating ranges (a few tens of centimeters). We vary the capturing distance between 10 and 50 cm and compute the spectral signature from the

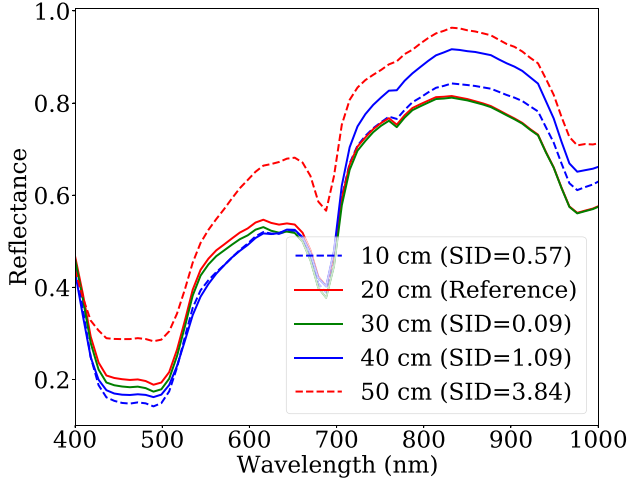


Fig. 19. Effect of capturing distance on the accuracy of spectral signatures computed from reconstructed bands.

reconstructed bands in each case. The accuracy of the spectral signatures is crucial for the system's operation, as significant deviations would lead to incorrect spectral analysis and ripeness assessment.

We objectively quantify the signature accuracy by computing the SID metric, which measures the similarity between the probability distributions of two signatures. We use the signature at 20 cm as the reference, as this is a typical distance for NIR cameras and produced the best results in our experiments. The results, shown in Fig. 19, indicate that signatures computed from capturing distances between 20–30 are fairly accurate and close to the reference signature. The accuracy drops outside of this range because the quality of the NIR image deteriorates, which in turn impacts the accuracy of the reconstructed bands. The strength of the NIR signal rapidly decreases at distances  $\geq 40$  cm, leading to inaccurate signatures, as shown in the figure. When the capturing distance is too small ( $\leq 10$  cm), the phone fails to capture all reflected NIR signals, negatively impacting accuracy.

**Total Run Time of RipeTrack on Phones:** We deployed RipeTrack on the Google Pixel 4XL phone and measured the total execution time, from capturing the RGB and NIR images to producing the final output on the screen. The average execution time is 191 milliseconds, which includes the two main spectral reconstruction and classification models, as well as other smaller tasks such as image alignment, scaling, and normalization.

**Complexity of the Reconstruction Model:** The reconstruction model of RipeTrack has a total of 293,356 parameters. The model took about 1 h and 45 minutes to train on the hyperspectral images dataset using a workstation with an NVIDIA Titan RTX GPU (24 GB memory), 32 GB main memory, and 3.60 GHz 16-core (Intel i9-9900 K) processor. The trained reconstruction model has a size of 8 MB. During inference, the reconstruction model uses about 3.1 GB of GPU memory. The average inference time on the Google Pixel 4XL phone is 0.11 seconds to reconstruct 68 bands.

**Complexity of the Classification Model:** The classification model has a total of 38,551,475 parameters, and it took about 10

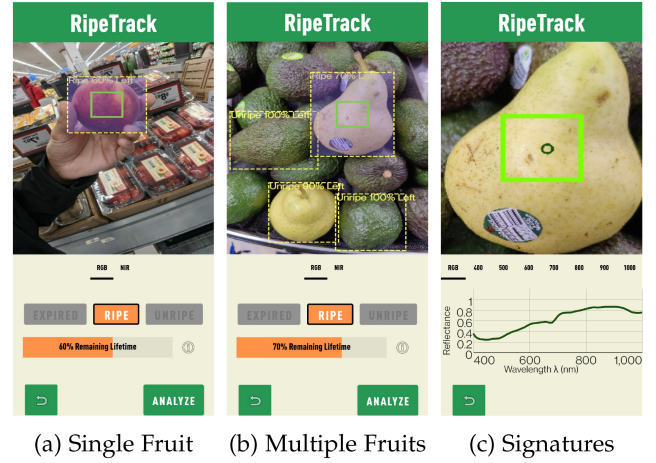


Fig. 20. Sample results from in-store testing. RipeTrack can analyze fruits in realistic environments on unmodified phones. It also provides a detailed spectral analysis.

TABLE IV  
DATASET CAPTURED IN FIVE GROCERY STORES UNDER DIVERSE AND REALISTIC ILLUMINATIONS AND FRUIT ARRANGEMENTS

| Fruit     | Count |
|-----------|-------|
| Pear      | 12    |
| Avocado   | 28    |
| Banana    | 32    |
| Mango     | 12    |
| Nectarine | 18    |
| Mixed     | 12    |
| Total     | 114   |

hours to train on the workstation mentioned above. The trained classification model uses 5.4 GB of GPU memory. The average inference time to output the ripeness level and remaining lifetime on the Google Pixel phone is 36 milliseconds.

**In-Store Testing:** We tested RipeTrack in five grocery stores that have different settings and illuminations. One of these stores is a small neighbourhood produce retailer and has its own farms in the city. The other four belong to major chains, e.g., Walmart Supercenter. We show sample screenshots of these tests in Fig. 20. RipeTrack implements object detection, which displays dotted boxes around all identified objects and fruits. The user can then remove any irrelevant object by tapping and holding it. When a user selects a fruit, RipeTrack analyzes a patch of  $64 \times 64$  pixels and displays the estimated ripeness level and remaining lifetime (Fig. 20(a)). As shown in Fig. 20(b), RipeTrack can analyze multiple different fruits at the same time. RipeTrack also allows interested users to visualize and inspect the spectral signatures and various bands (Fig. 20(c)).

In total, we collected 114 samples of five different fruits and their mixtures from five grocery stores, as summarized in Table IV. This dataset was collected by capturing RGB and NIR images by holding the phone at approximately 20–30 cm. This dataset is used for testing only; the models of RipeTrack have never seen it before. Since this is an *uncontrolled* environment and we cannot know the ground truth of fruit lifetime, we could only conduct a subjective analysis, or sanity check, of

the results produced by RipeTrack . Specifically, we opted to capture images of mostly unripe fruits or fruits at an early stage of ripening, based on our own intuition and visual inspection. Then, we assess the ripeness and remaining lifetime using RipeTrack . Overall, RipeTrack classified 96% of the samples as Unripe and the remaining 4% as Ripe. It also produced 80%–100% lifetime remaining for the samples. We believe the results are reasonably accurate and in accordance with grocery stores' tendency to sell fresh fruits to maintain customer satisfaction.

## IX. CONCLUSION

Accurately and easily estimating fruit ripeness reduces food waste, saves precious natural resources, and helps consumers and retailers lower costs. We presented a cost-effective approach that contributes to achieving this goal. We first showed that fruit ripeness can be assessed by spectral analysis in the visible and NIR (400–1000 nm) range using a hyperspectral camera. This is similar to the sensitivity range of CMOS sensors on phone cameras. However, phone cameras typically remove all signals beyond the visible range ( $>700$  nm) because they may damage image quality. We presented methods to obtain NIR signals and accurately reconstruct the spectrum in the entire 400–1000 nm range. We then presented RipeTrack, a mobile application that performs spectral analysis of fruits and predicts their ripeness level and remaining lifetime. Through extensive experimentation, we demonstrated that RipeTrack achieves high accuracy for various fruits and generates intuitive outputs for retailers and consumers. We also showed that RipeTrack can easily be extended to new fruits using transfer learning, and it is robust to diverse phones, illuminations, and capturing distances.

## REFERENCES

- [1] H. Forbes, T. Quested, and C. O'Connor, "Food waste index report 2021," United Nations Environment Programme, Tech. Rep., 2021. [Online]. Available: <https://www.unep.org/resources/report/unep-food-waste-index-report-2021>
- [2] D. Gunders, "Wasted: How America is losing up to 40 percent of its food from farm to fork to landfill (Second Edition)," Natural Resources Defense Council (NRDC), online, Tech. Rep., 2017. [Online]. Available: <https://www.nrdc.org/resources/wasted-how-america-losing-40-percent-its-food-farm-fork-landfill>
- [3] "The United States department of agriculture (USDA) economic research service (ERS): Estimates of food loss at the retail and consumer levels," 2024. [Online]. Available: <https://www.ers.usda.gov/data-products/food-availability-per-capita-data-system/food-loss/>
- [4] A. Batu, "Determination of acceptable firmness and colour values of tomatoes," *J. Food Eng.*, vol. 61, no. 3, pp. 471–475, 2004.
- [5] F. B. Abeles, P. W. Morgan, M. Saltveit Jr, and M. E. Saltveit Jr, *Ethylene in Plant Biology*. San Francisco, CA, USA: Academic, Jan., 1992.
- [6] "StrellaBiotech.com," 2024. [Online]. Available: <https://www.strellabiotech.com/>
- [7] S. Sohaib Ali et al., "Towards fruit maturity estimation using NIR spectroscopy," *Infrared Phys. Technol.*, vol. 111, 2020, Art. no. 103479.
- [8] S. S. Afzal, A. Kludze, S. Karmakar, R. Chandra, and Y. Ghasempour, "AgriTera: Accurate non-invasive fruit ripeness sensing via sub-terahertz wireless signals," in *Proc. ACM Conf. Mobile Comput. Netw.*, 2023, pp. 1–15.
- [9] S. Karmakar, A. Kludze, and Y. Ghasempour, "Meta-sticker: Sub-terahertz metamaterial stickers for non-invasive mobile food sensing," in *Proc. ACM Conf. Embedded Netw. Sensor Syst.*, 2023, pp. 335–348.
- [10] F. Shang, P. Yang, Y. Yan, and X. Y. Li, "LiqRay: Non-invasive and fine-grained liquid recognition system," in *Proc. ACM Conf. Mobile Comput. Netw.*, 2022, pp. 296–309.
- [11] U. Ha, J. Leng, A. Khaddaj, and F. Adib, "Food and liquid sensing in practical environments using RFIDs," in *Proc. USENIX Symp. Networked Syst. Des. Implementation*, 2020, pp. 1083–1100.
- [12] B. Sun, S. R. X. Tan, Z. Ren, M. C. Chan, and J. Han, "Detecting counterfeit liquid food products in a sealed bottle using a smart-phone camera," in *Proc. ACM Conf. Mobile Syst., Appl. Serv.*, 2022, pp. 42–55.
- [13] S. Yue and D. Katabi, "Liquid testing with your smartphone," in *Proc. ACM Conf. Mobile Syst., Appl., Serv.*, 2019, pp. 275–286.
- [14] "RipeTrack GitHub repository," 2024. [Online]. Available: <https://github.com/ShahzaibWaseem/RipeTrack>
- [15] OECD, "Guidelines on objective tests to determine quality of fruit and vegetables, dry and dried produce," OECD Publishing, Tech. Rep., 2018. [Online]. Available: <https://www.oecd.org/agriculture/fruit-vegetables/publications/guidelines-on-objective-tests.pdf>
- [16] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," *Artif. Intell. Agriculture*, vol. 7, pp. 44–57, 2023.
- [17] "Produce quality meter (F-750)," 2024. [Online]. Available: <https://felixinstruments.com/food-science-instruments/nir-spectroscopy/f-750-produce-quality-meter/>
- [18] O. O. Olarewaju, I. Bertling, and L. S. Magwaza, "Non-destructive evaluation of avocado fruit maturity using near infrared spectroscopy and PLS regression models," *Scientia Horticulturae*, vol. 199, pp. 229–236, 2016.
- [19] Y. Huang, K. Chen, Y. Huang, L. Wang, and K. Wu, "Vi-liquid: Unknown liquid identification with your smartphone vibration," in *Proc. ACM Conf. Mobile Comput. Netw.*, 2021, pp. 174–187.
- [20] J. Yun et al., "PowDew: Detecting counterfeit powdered food products using a commodity smartphone," in *Proc. ACM Conf. Mobile Syst., Appl. Serv.*, 2024, pp. 210–222.
- [21] H. Hu, Y. Zhu, B. Yang, H. Kang, S. Chen, and Q. Zhang, "Meat-Spec: Enabling ubiquitous meat fraud inspection through consumer-level spectral imaging," in *Proc. ACM Conf. Mobile Comput. Netw.*, 2024, pp. 861–874.
- [22] N. Sharma, M. S. Waseem, S. Mirzaei, and M. Hefeeda, "MobiSpectral: Hyperspectral imaging on mobile devices," in *Proc. ACM Conf. Mobile Comput. Netw.*, 2023, pp. 1–15.
- [23] Y. Cai et al., "MST : Multi-stage spectral-wise transformer for efficient spectral reconstruction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2022, pp. 745–755.
- [24] R. Pu, *Hyperspectral Remote Sensing: Fundamentals and Practices*. Evanston, IL, USA: Routledge, 2017.
- [25] J. Lammertyn, A. Peirs, J. De Baerdemaeker, and B. Nicolai, "Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment," *Postharvest Biol. Technol.*, vol. 18, pp. 121–132, Apr., 2000.
- [26] C. Sun, B. Aernouts, R. VanBeers, and W. Saeys, "Simulation of light propagation in citrus fruit using Monte Carlo multi-layered (MCML) method," *J. Food Eng.*, vol. 291, 2021, pp. 110225.
- [27] E. Z. D. Pratiwi, M. F. R. Pahlawan, D. N. Rahmi, H. Z. Amanah, and R. E. Masithoh, "Non-destructive evaluation of soluble solid content in fruits with various skin thicknesses using visible–shortwave near-infrared spectroscopy," *Open Agriculture*, vol. 8, no. 1, 2023, Art. no. 20220183.
- [28] B. Arad et al., "NTIRE 2022 spectral recovery challenge and data set," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2022, pp. 863–881.
- [29] J. Li, C. Wu, R. Song, Y. Li, and F. Liu, "Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops*, 2020, pp. 1894–1903.
- [30] "Raspberry pi cameras," 2024. [Online]. Available: <https://www.raspberrypi.com/documentation/accessories/camera.html>
- [31] Y. Monno, H. Teranaka, K. Yoshizaki, M. Tanaka, and M. Okutomi, "Single-sensor RGB-NIR imaging: High-quality system design and prototype implementation," *IEEE Sensors J.*, vol. 19, no. 2, pp. 497–507, Jan. 2019.
- [32] "ON Semiconductor's CMOS image sensor model AR0237 RGB-IR," 2024. [Online]. Available: <https://www.onsemi.com/products/sensors/image-sensors/AR0237>

- [33] A. Dosovitskiy et al., “An image is worth  $16 \times 16$  words: Transformers for image recognition at scale,” in *Proc. Int. Conf. Learn. Representations*, 2021, pp. 1–22.
- [34] A. Vaswani et al., “Attention is all you need,” in *Proc. Conf. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.
- [35] C.-I. Chang, “An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis,” *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1927–1932, Aug. 2000.
- [36] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. A. Ahmad, “A comparative study of loss functions for hyperspectral SISIR,” in *Proc. Eur. Signal Process. Conf.*, 2022, pp. 484–487.
- [37] M. Afifi and M. S. Brown, “Deep white-balance editing,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Seattle, WA, USA, 2020, pp. 1394–1403.
- [38] M. Liu, J. Pirrello, C. Chervin, J.-P. Roustan, and M. Bouzayen, “Ethylene control of fruit ripening: Revisiting the complex network of transcriptional regulation,” *Plant Physiol.*, vol. 169, no. 4, pp. 2380–2390, 2015.
- [39] “Ethylene detector (FD-90A-C2H4),” 2024. [Online]. Available: <https://www.forensicsdetectors.com/blogs/articles/ethylene-gas-detector-produce>
- [40] V. Paul, R. Pandey, and G. C. Srivastava, “The fading distinctions between classical patterns of ripening in climacteric and non-climacteric fruit and the ubiquity of ethylene—an overview,” *J. Food Sci. Technol.*, vol. 49, no. 1, pp. 1–21, 2012.
- [41] D. H. Foster and K. Amano, “Hyperspectral imaging in color vision research: Tutorial,” *J. Opt. Soc. Amer. A*, vol. 36, no. 4, Art. no. 606, 2019.
- [42] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [43] Z. Wang, A. C. Bovik, and E. P. Simoncelli, “Structural approaches to image quality assessment,” in *Proc. Handbook Image Video Process.*, 2005, pp. 961–974.



**Muhammad Shahzaib Waseem** received the BSc degree in computer science from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2020, and the MSc degree in computing science from Simon Fraser University (SFU), Burnaby, BC, Canada, in 2024. His research interests include computer vision, hyperspectral imaging, and machine learning.



**Neha Sharma** received the BTech degree in computer science from the Motilal Nehru National Institute of Technology (MNNIT), Allahabad, India, in 2015. She is currently working toward the PhD degree in the School of Computing Science, Simon Fraser University (SFU), Burnaby, BC, Canada. Her research interests include multimedia systems, mobile sensing, hyperspectral imaging, and machine learning.



**Mohamed Hefeeda** (Senior Member, IEEE) received the BSc and MSc degrees from Mansoura University, Mansoura, Egypt, in 1994 and 1997, respectively, and the PhD degree from Purdue University, West Lafayette, IN, USA, in 2004. He is a professor in the School of Computing Science, Simon Fraser University (SFU), Canada, where he leads the Network Systems Lab. He was named an ACM Distinguished Member, in 2025. Dr. Hefeeda's research interests include multimedia systems and computer networks. In 2011, he was awarded one of the prestigious NSERC

Discovery Accelerator Supplements (DAS), which are granted to a select group of distinguished researchers from all Science and Engineering disciplines, in Canada. Dr. Hefeeda's research received multiple paper awards, published in reputable journals and conferences, and resulted in ten granted patents, a start-up company, and multiple technology transfers to major companies. Dr. Hefeeda served as the director of the School of Computing Science, SFU between 2018 and 2023, where he led a major faculty renewal and expansion process. Under his leadership, the School hired 25 faculty members, introduced a new Professional master's program in Cybersecurity, and substantially increased its graduate and undergraduate enrollments. He served on the editorial boards of premier journals such as the *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*, where he was named the best associate editor, in 2014. He also served on the organization committees and/or co/chaired several international conferences such as ACM Multimedia, ACM Multimedia Systems, IEEE Infocom, and IEEE ICME.