RipeTrack: Assessing Fruit Ripeness and Remaining Lifetime using Smartphones

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

Abstract—Several studies showed that a significant fraction of fresh fruits are discarded at the retail and consumer levels, wasting precious resources, polluting the environment, and increasing 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 utilizes the sensing capabilities of phones and machine learning models to analyze the optical properties of fruits in different 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 95%. 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—Mobile applications, hyperspectral imaging, fruit ripening

1 INTRODUCTION

Food 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 selling such fruits, leading to significant food waste and loss of revenues. Similarly, consumers may purchase fruits that are not suitable for their use (either close to expiration or not sufficiently ripe yet), likely discarding such fruits.

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 provide higher accuracy than manual inspection, they require special hardware setups and are way 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 have quite diverse ripening patterns and lifetimes, and a general solution should account for such 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 complicate 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 stages without damaging them. It then maps these properties to easy-to-understand categories by consumers

The authors are with the School of Computing Science, Simon Fraser University, Burnaby, BC, Canada.

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 §4. 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 propose three alternative solutions to conduct spectral analysis in the 400–1000 nm range, which enables assessing fruit ripeness and remaining lifetime on smartphones in §5. One of these solutions does not require any hardware changes, while the other two are easily realizable.
- 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 §6.
- We implement a mobile application, RipeTrack, to demonstrate the practicality of the proposed approach in §7.
- We conduct an extensive evaluation study to analyze the accuracy, robustness, and extensibility of RipeTrack in §8. 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] use sub-tera Hertz waves that 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 [14] tests water contamination. We summarize the related works in §2. The code and datasets of this work are **open source** [15].

2 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 [16], [17], [7], 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., [18], 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 substages, 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 [16] 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. [17] summarize automated approaches that utilize machine learning to assess ripeness using images captured by regular RGB cameras. However, as demonstrated in §4, the external colors and features of some fruits may not accurately reflect the chemical changes happening inside the 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. [19] 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 [12], [20], [14], [11], [10], [13]. LiquidHash [12] models the motion of air bubbles inside bottles to detect adulteration in liquids. CapCam [14] estimates the surface tension of liquids to identify alcohol concentration and water contamination levels. Vi-Liquid [20] identifies liquids by measuring their viscosity coefficients using phone accelerometers. RF-EATS [11] and LiqRay [10] use RFID tags to differentiate various liquids. MobiSpectral [13] 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 [21]. We compare the proposed reconstruction model versus the one in MobiSpectral and show that our model is more efficient. For example, in the inference mode, it runs 30X faster and requires 2.4X less memory compared to MobiSpectral, which is crucial for mobile platforms with limited resources. Our model also produces better reconstruction results, as shown in §8.3.

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.

3 Problem Definition and Challenges

The problem addressed in this paper is how to determine the ripeness level and remaining lifetime of 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 [22]. 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 §4, we first conduct an experimental study to investigate the feasibility of assessing fruit ripeness using spectral analysis, utilizing a hyperspectral camera that provides detailed information across more than 200 bands.

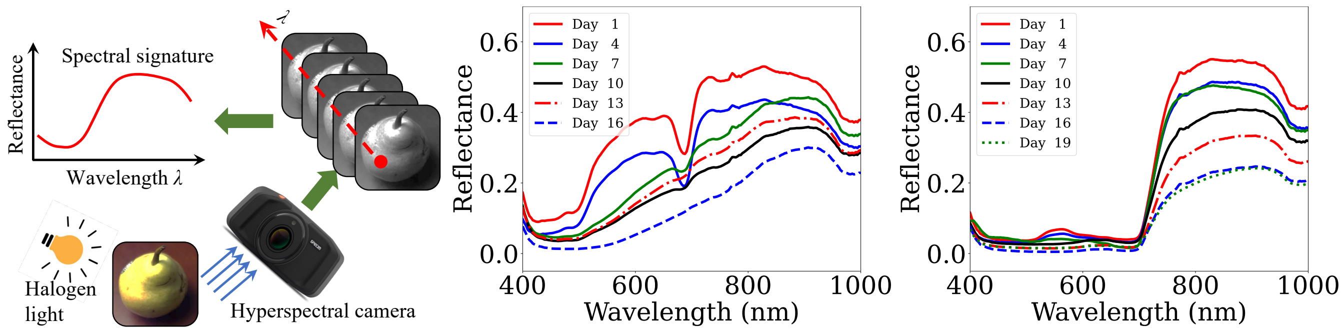
Hyperspectral cameras are expensive (tens of thousands of dollars) and require strict illumination (halogen sources) that are hard to achieve in environments such as grocery stores and homes. In §5, we propose a method to conduct spectral analysis on phones. This method uses various signals captured 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 §6, 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 980nm) 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 §5.3, we present methods to handle this diversity and improve the robustness of the proposed system.

4 Tracking Internal Changes using Hyperspectral Cameras

Prior works have shown that spectral analysis in different ranges of the spectrum can identify organic materials. For example, soluble solids, e.g., sugars, can be observed in the 750–1100 nm range [23], oil content in avocados can be measured in the 2200–2400 nm range [19], water content can be discerned in the 960–980 nm range [24], and the breaking down of Chlorophyll into pigmentation can be tracked in the visible light (400–700 nm) range [16]. 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



(a) Setup (b) Pear (Bartlett) (c) Avocado (Organic)

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

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.

Figure 1a 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 512×512 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 §8.1, 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 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 each fruit, we compute a signature for each day of its lifetime. We present representative signatures of Pear (Bartlett) and Avocado (Organic) in Figure 1; other fruits exhibit similar patterns. These two fruits are quite different 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 Figure 1b. As pears ripen, their exterior color gradually changes since chlorophyll (greenish color) breaks down into new pigmentation (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 as a result of 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 Figure 1c, 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 show 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 done with 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 [23], [19], [24], [16], 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.

5 Tracking Internal Changes on Phones

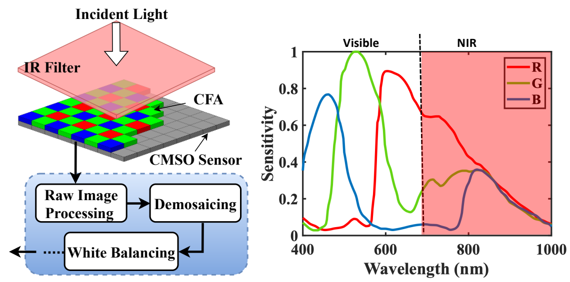
5.1 Overview

The spectral analysis in §4 was conducted with an expensive (30+K 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 challenging problem as discussed below.

Smartphone cameras are much simpler than hyperspectral cameras. As illustrated in Figure 2a, a smartphone camera uses a color filter array (CFA) that allows light in the Red (R), Green (G), and Blue (B) wavelength bands to be captured by the 2-dimensional CMOS sensor on the camera. The most common CFA is the Bayer pattern shown in the figure, where more pixels are allocated to the green band than the blue and red bands since the human visual system is more sensitive to the green color. Each pixel (photo diode) 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, other steps, e.g., white balancing, color correction, and compression, are performed to produce the output RGB image. We note 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.

Because of their relatively simple design, smartphone cameras produce coarse-grain information (only three bands) in the visible range, which is not sufficient for spectral analysis as it requires many bands to create spectral signatures. Further, smartphone cameras produce no information in the NIR range, which is essential for fruit ripening analysis because signals in the NIR range can penetrate the fruit surface and reveal changes happening inside the fruit, as discussed in §4.

To address these challenges and enable conducting spectral analysis on smartphones, we propose *upscaling* the few captured bands by smartphone cameras to many bands. This is sometimes referred to as spectral reconstruction in the literature [25], [26], [21]. In the **Supplementary Materials** (§A.1), we analyze the state-of-the-art reconstruction model [21] for the suitability of conducting spectral analysis of fruits. Our analysis shows that this model provides low-quality reconstructed bands in the NIR range, resulting in significant errors compromising the accuracy of spectral signatures created from such bands. One of the main reasons behind this poor performance is the lack of any NIR signals in the input, which makes the model hallucinate bands in the NIR range. To address this problem, we propose three possible solutions to obtain NIR signals in §5.2. Then, in §5.3, 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.



(a) Camera pipeline (b) Spectral sensitivity

Figure 2: Smartphone cameras and their sensitivity. The IR filter removes all signals beyond 700 nm.

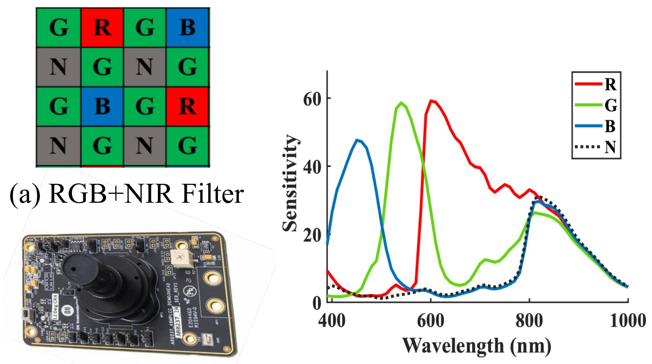
5.2 Acquiring NIR Signals on Smartphones

As mentioned before, NIR signals are crucial to accurately reconstruct bands across the whole 400-1000 range. We present three practical solutions to obtain NIR signals on smartphones: (i) Using Full Spectrum RGB Cameras, (ii) Designing a custom color filter array, and (iii) Using the NIR camera on modern smartphones. We analyze the pros and cons of each solution, and we experimentally evaluate their performance in §8.2.

Using Full Spectrum RGB Cameras. Some commercial cameras, e.g., Raspberry PI Camera Module 3 [27], come *without* IR filters. They are usually referred to as full spectrum cameras as they capture signals in most of the 400–1000 nm sensitivity range of CMOS sensors. Applications of such cameras include infrared photography and surveillance. An equivalent approach is to remove the IR filter from regular RGB cameras. The IR filter is typically a thin film attached to the camera sensor, which can be removed using a relatively easy procedure [28]. To analyze the effectiveness of this approach, we removed the IR filter from a smartphone (Model: Google Nexus 5X). We plot in Figure 2b the sensitivity of the CMOS sensor. The availability of signals in the NIR range, even if mixed with the RGB channels, improves the reconstruction model as shown in §8.2. We note that most CMOS sensors in the market have a sensitivity pattern similar to the one shown in the *left* part of Figure 2b, and all signals beyond 700 nm are truncated by the IR filter.

The advantages of this approach to obtain NIR signals are the simplicity and existence of commercial cameras without IR filters [27]. However, cameras without IR filters damage the quality of regular images. Thus, this solution is more suitable for imaging systems designed for special tasks such as quality and grading inspection devices.

Designing Custom Color Filter Array. Cameras without IR filters allow the CMOS sensor to *implicitly* capture NIR signals mixed with the RGB bands (Figure 2b). This provides limited NIR information and damages the visual quality of RGB images. We consider another approach that preserves



(a) RGB+NIR Filter (b) Sensor Board (c) Sensitivity

Figure 3: The camera module with custom color filter array used in our experiments.

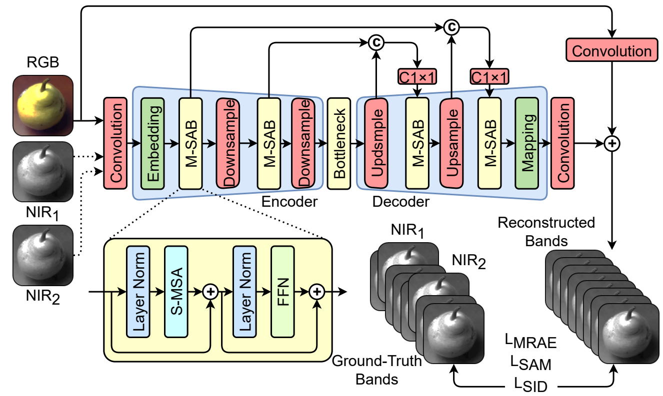


Figure 4: Architecture of the proposed spectral reconstruction model.

visual quality and provides a separate NIR channel. This is done by replacing the traditional Bayer color filter array (Figure 2a) with a custom-designed one with an explicit filter for the NIR channel.

Designing a custom filter array is a complex research problem in its own right since there are numerous possible filter patterns, and the performance of each pattern depends on several factors, including the CMOS sensor sensitivity and the processing steps performed on the output of the sensor. This is further complicated by the fact that not all theoretically possible filter patterns can be manufactured in practice. Some of them, for example, are too dense and may result in cross-band interference. Notice that the color filter array is itched on the CMOS sensor, unlike the IR filter, which can be removed.

Monno et al. [29] conduct a comprehensive performance analysis of various color filter arrays. They demonstrate that the 4×4 filter pattern shown in §5.2.a results in the best overall performance in terms of obtaining an explicit NIR band with minimal damage to the RGB bands. This design, however, was analyzed using only *synthetic* data, not real camera sensors. We found and purchased a commercial camera sensor that has a similar filter pattern, which is the CMOS Image Sensor Model AR0237 RGB-IR from ON Semiconductor [30]. This sensor and its sensitivity are shown in §5.2.b and §5.2.c, respectively. The sensitivity was obtained from the manufacturer's data sheets. We analyze the potential of reconstructing all bands in the 400-1000 nm range using the four RGB+NIR bands offered by this filter and compare the results against the ground truth in §8.2.

The advantages of using custom filters include better spectral reconstruction and maintaining the quality of RGB images. The disadvantage is the cost of manufacturing imaging sensors with custom filters. This solution can be useful for designing future smartphones and specialized imaging systems for quality analysis and inspection.

Using RGB and NIR Cameras. 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.

To enable spectral reconstruction on *current* smartphones, we propose using their NIR cameras. This does not require any hardware changes, which is a major advantage. However, it needs to control two cameras to capture images right after each other. It also needs to resize and align the images, as the two cameras produce different resolutions and are physically apart on the phone. We analyze the accuracy of the reconstruction model when the input is RGB and NIR images in §8.2.

5.3 Spectral Reconstruction Model

The proposed spectral reconstruction model is shown in Figure 4, which is designed using vision transformers [31] similar to recent works, e.g., [21]. Compared to prior works, however, our model considers the NIR band as an extra input, adds 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 smartphones.

Vision transformers can efficiently learn correlations in the input data through a mechanism known as self attention [32]. 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 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 [21], which computes the attention across spectral bands.

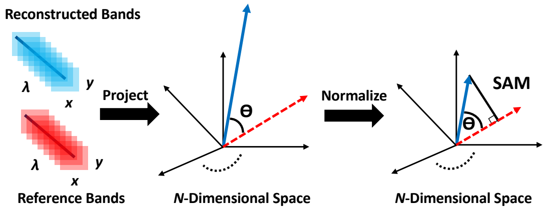


Figure 5: Illustration of the SAM loss function.

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 reconstructed and ground truth bands. It strives to ensure the accuracy of the reconstructed bands in the spatial domain, and it is computed as:

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

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 [22]. Figure 5 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:

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

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 [33]. SID first transforms the spectral bands into probability distributions, and it then calculates the difference between them, as illustrated in Figure 6. SID is computed as:

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

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:

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

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 [34].

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.

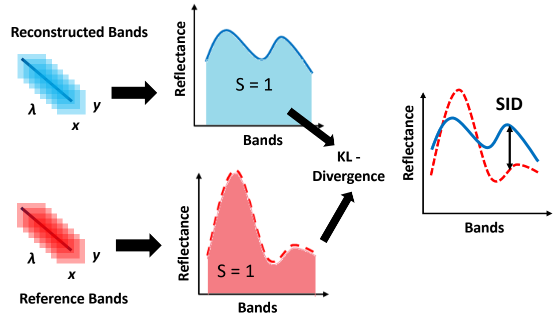


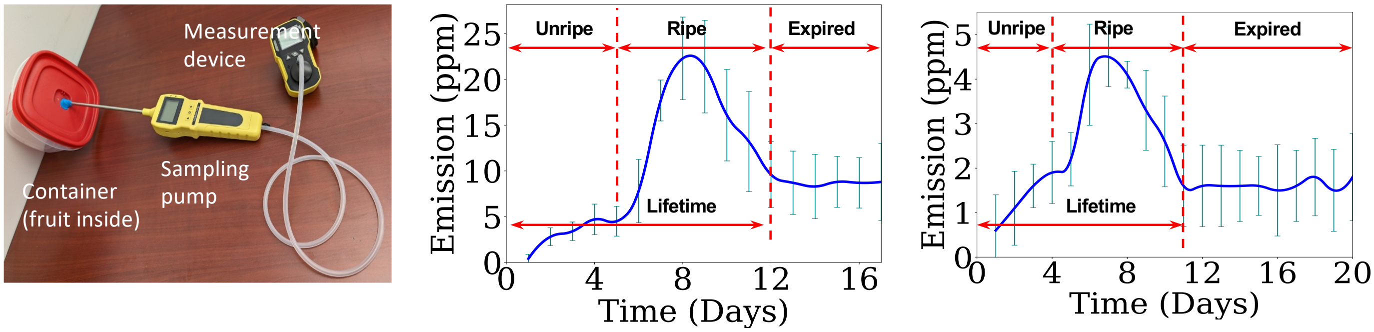
Figure 6: Illustration of the SID loss function.

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 Figure 4, 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_1), (RGB, NIR_2), ..., (RGB, NIR_L) 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 illumination of the scene, in addition to the 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 use an image normalization approach similar to [13], where we implement a deep-learning model that maps an input RGB image to a common representation regardless of the characteristics of the camera and scene illumination. All RGB images are first normalized before being used in the reconstruction model.

6 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



(a) Setup (b) Pear (Bartlett) (c) Avocado (Organic)

Figure 7: Analysis of the ethylene emission rate (in parts per million or ppm) for two fruits over their lifetime.

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 [35].

To develop our labeling method, we conduct experiments to analyze the ethylene emission rate for different fruits. Our experimental setup is shown in Figure 7a, 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 [36], 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 Figure 7b and Figure 7c, 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.

Prior food science research, e.g., [37], 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 Figure 7b and Figure 7c 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 §7.2. During inference, which is done on smartphones, ethylene measurement is *not* performed; we only use images.

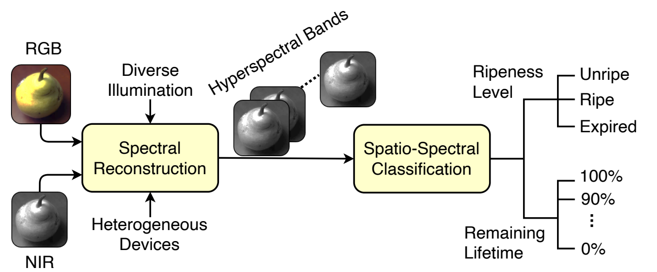


Figure 8: Overview of the proposed system to assess fruit ripeness and remaining lifetime on smartphones.

Alternatives. 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 chart charts [17], [16], can be used to provide approximate labeling.

7 End-to-End System and Mobile App

7.1 System Overview and Operation

Figure 8 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 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

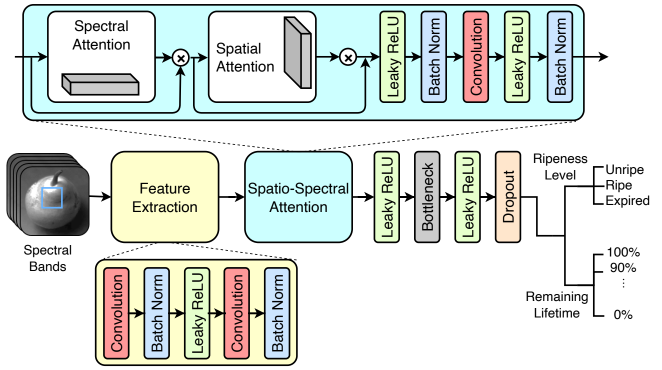


Figure 9: Design of the proposed spatio-spectral classifier for fruit ripeness level and remaining lifetime.

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.

7.2 Design of the Spatio-Spectral Classifier

We propose a classification model that considers both the spatial and spectral characteristics of the input hyperspectral 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 Figure 9. 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 given to a Spatio-Spectral Attention module, which has two attention blocks. The first captures the context 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.

7.3 Mobile App

We have developed an Android application as a proof of concept; sample screenshots are given in Figure 21, and the code can be found at [15]. The application is written in Kotlin and complied using Gradle version 8.0. The image capturing modules of the application are built on top of the Android's 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 and ported to the Android platform using the PyTorch's JIT Trace. The trained models are then integrated with the 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 right after it captures 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 can store intermediary data, e.g., reconstructed bands, for debugging and further analysis.

7.4 Limitations and Extensions

RipeTrack requires NIR signals, and we presented three solutions to acquire such signals. One of these solutions provides good accuracy without modifying 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 hyperspectral imaging dataset, which is open-source [15], 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, or they can be approximated using manual methods such as comparing fruit colors versus standard charts and/or performing firmness tests.

8 Evaluation

We first describe our setup and datasets in §8.1. Then, we demonstrate the accuracy of the proposed three methods for conducting spectral analysis on phones in §8.2. In §8.3, 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 §8.4. In §8.5, 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 [15], with details to reproduce our results.

8.1 Experimental Setup

Testbed. Figure 10 shows our testbed, which has:

- ***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 3nm. The spatial resolution of each band is 512×512 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.

Figure 10: The testbed used in our experiments.

- ***Camera Module with Custom RGB+NIR CFA:*** Used to evaluate the spectral reconstruction model. The model is ON Semiconductor AR0237 [30].
- ***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 [36], 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 1, 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 Figure 7. 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 §6.

	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	

Table 1: Summary of the fruit ripening dataset.

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 1. 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 390K 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 800×600 and 640×480 pixels for the RGB and NIR cameras, respectively, whereas OnePlus has resolutions of 4032×3024 and 2592×1944 pixels. We scale all RGB and NIR images to 640×480 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.

8.2 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

	SAM ↓	SID ↓	PSNR ↑	SSIM ↑
No IR	0.1029	0.0378	31.2	0.9460
Custom Filter	0.0620	0.0073	39.4	0.9856
RGB + NIR	0.0798	0.0132	34.0	0.9742

Table 2: Performance of the reconstruction model. Average metrics across all fruits are presented.

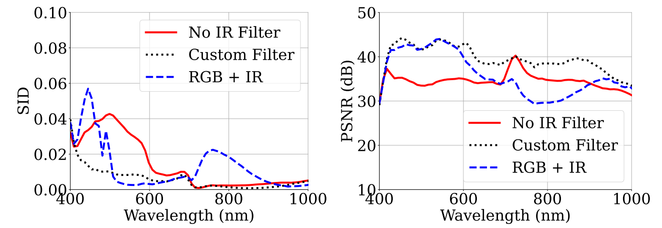
the reconstructed bands by comparing them against the ground-truth ones captured by the hyperspectral camera.

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 is three RGB bands when the IR filter is removed from the camera. The second version takes as input the four bands produced by the custom color filter array in §5.2, and it is referred to as Custom Filter. The third version represents the case of unmodified phones, and it takes 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 1 in this section, and we keep the others for later testing the extensibility of our model.

We measure the accuracy using six performance metrics commonly used in the literature [22], [13], 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 §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 2, where we show only averages and four metrics. Details of individual fruits and all metrics with confidence intervals are given §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 SAM and SID error metrics are close to zero.

To shed some light on the relative performance of the three versions, we plot their accuracy across individual bands in Figure 11 for two representative metrics. The results in Table 2 and Figure 11 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 damaged because of 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



(a) SID (b) PSNR

Figure 11: Band-wise performance analysis of the reconstruction model with different inputs.

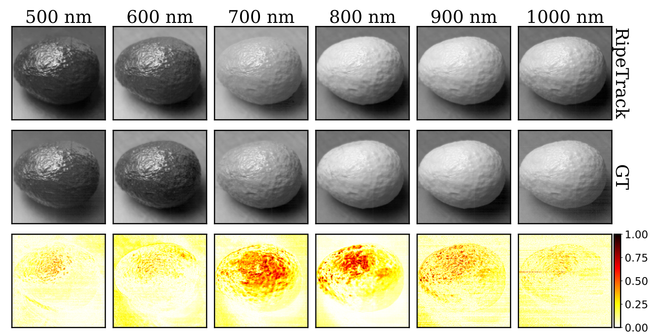


Figure 12: Visual comparison of the reconstructed bands and the ground truth (GT) ones; the bottom row shows the absolute errors between them as heat maps.

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

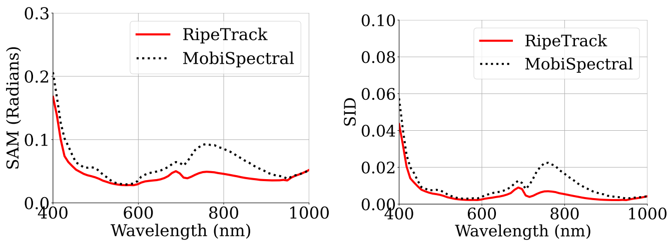
Finally, we show a few samples demonstrating the quality of the reconstructed bands using RGB and NIR images in Figure 12. The figure also shows the difference between each reconstructed band and its corresponding ground one as a heat map.

8.3 Comparison against State-of-the-Art

The proposed spectral reconstruction model in this paper improves on our prior work, MobiSpectral [13], by adding multiple loss functions to improve 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 [21].

We compare the performance of the proposed reconstruction model against MobiSpectral using the fruit ripening described in §8.1. In both cases, we use RGB and NIR images. We present sample results in Figure 13 for the two most important metrics: SAM and SID. 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 successively running both reconstruction



(a) SAM (b) SID

Figure 13: Performance of the proposed reconstruction model in RipeTrack versus state of the art (MobiSpectral).

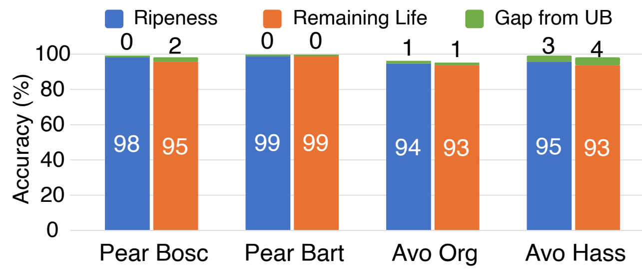


Figure 14: The accuracy of predicting fruit ripeness and remaining lifetime using images captured by phones.

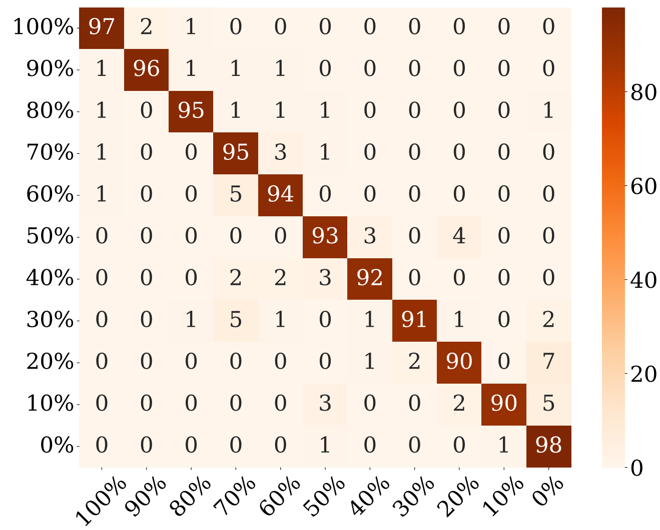


Figure 15: Accuracy of predicting individual classes of the remaining lifetime. Rows: predicted; Columns: actual.

models on the same workstation (specs are given in §8.5). The results are summarized in Table 3, which shows that the proposed reconstruction model is more efficient than MobiSpectral. For 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 about 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.

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

Table 3: Computational complexity of the reconstruction model in RipeTrack and state of the art (MobiSpectral).

8.4 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. The RGB and NIR images in the training partition are first upscaled 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 Figure 14, we summarize the average accuracy for 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 Figure 14 show, RipeTrack achieves high accuracy for all considered fruits. Specifically, an accuracy of at least 94% and 93% is observed for the 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 accuracy of assessing avocado ripeness is slightly lower than pear's, 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 Figure 15 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 the high accuracy across all classes, with classes at both ends of the lifetime achieving a 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 1. We extend RipeTrack to the other three fruits (banana, nectarine, and mango) by fine-tuning its reconstruction and classification models using transfer learning on a few samples from each fruit. We report the average accuracy in Figure 16, which confirms the extensibility and accuracy of RipeTrack.

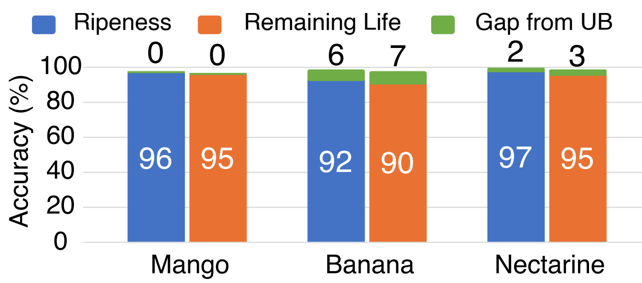


Figure 16: Extensibility of RipeTrack to other fruits using transfer learning by fine-tuning the model on a few images.

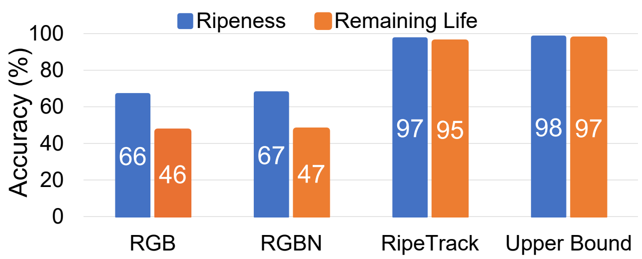


Figure 17: Ablation study: Performance impact of various components of RipeTrack.

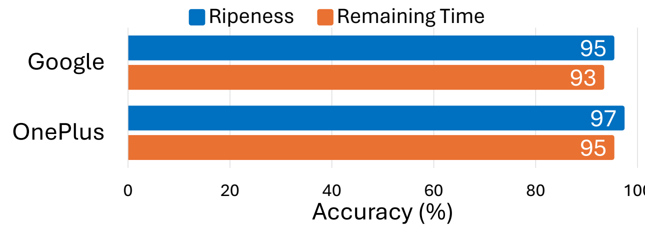


Figure 18: Robustness to phone diversity: RipeTrack produces accurate results on different phones.

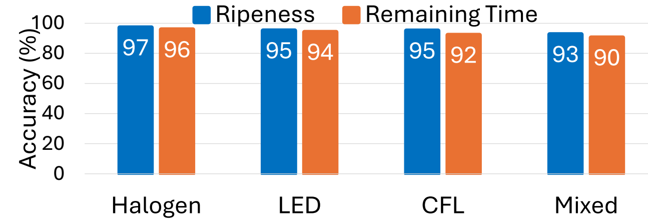


Figure 19: Robustness to illumination diversity: RipeTrack achieves high accuracy under different illuminations.

8.5 System Analysis and In-Store Testing

Ablation Study. We analyze the performance impact of different components of RipeTrack. Specifically, we evaluate the accuracy when we only use RGB images to assess fruit ripeness. Then, we add NIR images but with no spectral reconstruction. That is, the input to the classification model is a pair of RGB and NIR images. Then, we perform spectral reconstruction from RGB+NIR images and use the reconstructed bands as input to the classification model. The results of this experiment are shown in Figure 17, 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 the NIR and RGB bands to reconstruct the whole spectrum, providing rich information to the model.

Robustness to Practical Illuminations. Our mobile images dataset is captured with diverse illuminations: LED, Fluorescent (CFL), and mixtures of these sources and sunlight (referred to as Mixed). We assess the classification accuracy for each illumination source. The results in Figure 19 show that the image normalization method of RipeTrack mitigates the differences in illuminations, and the type of illumination does not impact the accuracy. The accuracy is the highest when using halogen sources, because they emit power across the whole spectrum, helping the phone to 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 to Diverse Phones. We separate the images captured by our two phones (Google Pixel and OnePlus) and compute the classification accuracy for each group. Our results in Figure 18 show insignificant differences in the accuracy achieved by RipeTrack for both phones. 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 reconstructed bands in each case. The accuracy of the spectral signatures is critical for the operation of the whole system, as significant deviations in them would lead to incorrect spectral analysis and ripeness assessment. The results, shown in Figure 20, indicate that signatures computed from capturing distances between 20–40 cm are fairly accurate and close to each other. The accuracy drops outside of this range because the quality of the NIR image suffers, which impacts the accuracy of the reconstructed bands. The strength of the NIR signal rapidly decreases at distances $\geq$ 50cm, leading to inaccurate signatures as shown in the figure. When the capturing

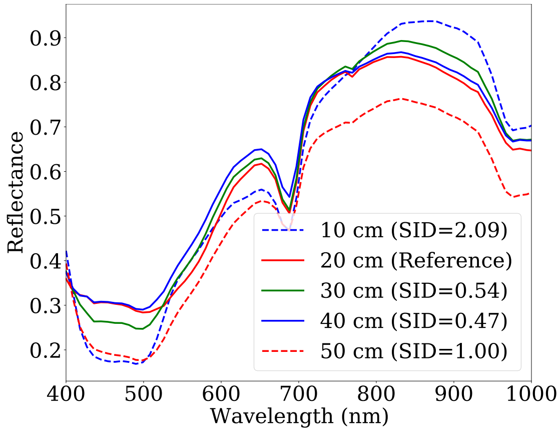
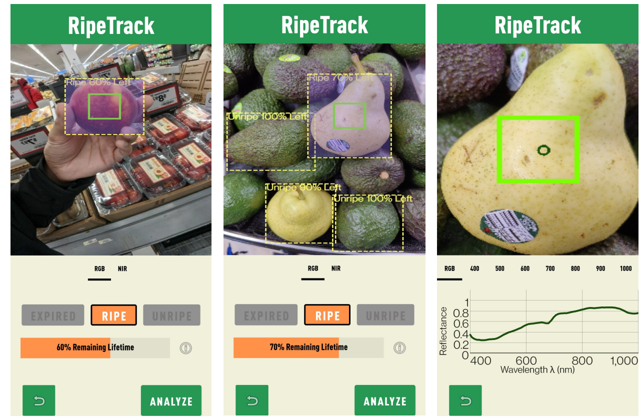


Figure 20: Effect of capturing distance on the accuracy of spectral signatures computed from reconstructed bands.



(a) Single Fruit (b) Multiple Fruits (c) Signatures

Figure 21: Sample results from in-store testing. RipeTrack can analyze fruits in realistic environments on unmodified phones. It also provides a detailed spectral analysis.

distance is too small ($\leq$ 10 cm), the phone does not capture all reflected NIR signals, negatively impacting the accuracy.

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, because this is a typical distance for NIR cameras, and it produced the best results in our experiments. The results are given in the legend of Figure 20, which show that the SID for distances $\geq$ 50 cm and $\leq$ 10 cm are quite high, indicating the large discrepancies between signatures at these distances and the reference one.

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 hour 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 i9-9900K) 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 sec 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 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 msec.

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 it 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 Figure 21. RipeTrack implements object detection, which displays dotted boxes around all identified objects, fruits and otherwise. The user can then remove any irrelevant object by tapping and holding it. When a user selects a fruit, RipeTrack analyzes a patch of 64×64 pixels and displays the estimated ripeness level and remaining lifetime (Figure 21a). As shown in Figure 21b, 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 (Figure 21c).

In total, we collected 114 samples of five different fruits and mixtures of fruits from the five grocery stores, as summarized in Table 4. This dataset was collected by capturing RGB and NIR images by holding the phone at approximately 20–40 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 early in their ripening stage, 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.

9 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

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

Table 4: Dataset captured in five grocery stores under diverse and realistic illuminations and fruit arrangements.

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 proposed 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 showed that RipeTrack yields high accuracy for different fruits and produces 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., 8 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," *Journal of Food Engineering*, 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 Diego, California: Academic Press, 1 1992.

[6] "StrellaBiotech.com," 2024. [Online]. Available: https://www.strellabiotech.com/

[7] S. Sohaib Ali Shah, A. Zeb, W. S. Qureshi, M. Arslan, A. Ullah Malik, W. Alasmary, and E. Alanazi, "Towards fruit maturity estimation using NIR spectroscopy," *Infrared Physics and Technology*, vol. 111, 12 2020.

[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 *Proceedings of ACM Conference on Mobile Computing and Networking (MobiCom'23)*, 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 *Proceedings of ACM Conference on Embedded Networked Sensor Systems (SenSys'23)*, 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 *Proceedings of ACM Conference on Mobile Computing and Networking (MobiCom'22)*, 10 2022, pp. 296–309.

[11] Unsoo Ha, Junshan Leng, Alaa Khaddaj, and Fadel Adib, "Food and Liquid Sensing in Practical Environments using RFIDs," in *Proceedings of USENIX Symposium on Networked Systems Design and Implementation (NSDI'20)*, 2020.

[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 smartphone camera," in *Proceedings of ACM Conference on Mobile Systems, Applications and Services (MobiSys'22)*, 6 2022, pp. 42–55.

[13] N. Sharma, M. S. Waseem, S. Mirzaei, and M. Hefeeda, "MobiSpectral: Hyperspectral Imaging on Mobile Devices," in *Proceedings of ACM Conference on Mobile Computing and Networking (MobiCom'23)*, ser. ACM MobiCom '23, 2023.

[14] S. Yue and D. Katabi, "Liquid testing with your smartphone," in *Proceedings of ACM Conference on Mobile Systems, Applications, and Services (MobiSys'19)*, 2019, pp. 275–286.

[15] "RipeTrack GitHub Repository ," 2024. [Online]. Available: https://github.com/ShahzaibWaseem/RipeTrack

[16] OECD, "Guidelines on objective tests to determine quality of fruit and vegetables, dry and dried produce," Organisation for Economic Co-operation and Development (OECD), Tech. Rep., 10 2018. [Online]. Available: https://www.oecd.org/agriculture/fruit-vegetables/publications/guidelines-on-objective-tests.pdf

[17] M. Rizzo, M. Marcuzzo, A. Zangari, A. Gasparetto, and A. Albarelli, "Fruit ripeness classification: A survey," *Artificial Intelligence in Agriculture*, vol. 7, pp. 44–57, 2023.

[18] "Produce Quality Meter (F-750)," 2024. [Online]. Available: https://felixinstruments.com/food-science-instruments/nir-spectroscopy/f-750-produce-quality-meter/

[19] 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.

[20] Y. Huang, K. Chen, Y. Huang, L. Wang, and K. Wu, "Vi-liquid: Unknown liquid identification with your smartphone vibration," in *Proceedings of ACM Conference on Mobile Computing and Networking (MobiCom'21)*, 2021, pp. 174–187.

[21] Y. Cai, J. Lin, Z. Lin, H. Wang, Y. Zhang, H. Pfister, R. Timofte, and L. Van Gool, "MST++: Multi-Stage Spectral-Wise Transformer for Efficient Spectral Reconstruction," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22) Workshops*, 2022, pp. 745–755.

[22] R. Pu, *Hyperspectral Remote Sensing: Fundamentals and Practices*. Routledge, 2017.

[23] C. Sun, B. Aernouts, R. Van Beers, and W. Saeys, "Simulation of light propagation in citrus fruit using monte carlo multi-layered (MCML) method," *Journal of Food Engineering*, vol. 291, 2 2021.

[24] 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.

[25] B. Arad and et al., "NTIRE 2022 Spectral Recovery Challenge and Data Set," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'22) Workshops*, 6 2022, pp. 863–881.

[26] 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 *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'20) Workshops*, 6 2020.

[27] "Raspberry Pi Cameras," 2024. [Online]. Available: https://www.raspberrypi.com/documentation/accessories/camera.html

[28] "Removal of IR Filter," 2024. [Online]. Available: https://www.raspberrypi.com/documentation/accessories/camera.html#filter-removal

[29] 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 Journal*, vol. 19, no. 2, pp. 497–507, 2019.

[30] "ON Semiconductor's CMOS Image Sensor Model AR0237 RGB-IR," 2024. [Online]. Available: https://www.onsemi.com/products/sensors/image-sensors/AR0237

[31] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in *Proceedings of International Conference on Learning Representations (ICLR'21)*, 2021.

[32] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is All you Need," in *Proceedings of Conference on Neural Information Processing Systems (NIPS'17)*, vol. 30, 2017.

[33] C.-I. Chang, "An information-theoretic approach to spectral variability, similarity, and discrimination for hyperspectral image analysis," *IEEE Transactions on Information Theory*, vol. 46, no. 5, pp. 1927–1932, 2000.

[34] N. Aburaed, M. Q. Alkhatib, S. Marshall, J. Zabalza, and H. A. Ahmad, "A Comparative Study of Loss Functions for Hyperspectral SISR," in *Proceedings of European Signal Processing Conference (EUSIPCO'22)*, 2022, pp. 484–487.

[35] 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 Physiology*, vol. 169, no. 4, pp. 2380–2390, 10 2015.

[36] "Ethylene Detector (FD-90A-C2H4)," 2024. [Online]. Available: https://www.forensicsdetectors.com/blogs/articles/ethylene-gas-detector-produce

[37] 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," *Journal of Food Science and Technology*, vol. 49, no. 1, pp. 1–21, 2 2012.

[38] D. H. Foster and K. Amano, "Hyperspectral imaging in color vision research: tutorial," *Journal of the Optical Society of America A*, vol. 36, no. 4, p. 606, 4 2019.

[39] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 4 2004.

[40] Z. Wang, A. C. Bovik, and E. P. Simoncelli, "Structural Approaches to Image Quality Assessment," *Handbook of Image and Video Processing, Second Edition*, pp. 961–974, 1 2005.

Muhammad Shahzaib Waseem received the B.Sc. degree in Computer Science from the National University of Sciences and Technology (NUST), Islamabad, Pakistan, in 2020, and the M.Sc. 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 B.Tech. degree in Computer Science from Motilal Nehru National Institute of Technology (MNNIT), Allahabad, India, in 2015. She is a PhD candidate in the School of Computing Science at 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 B.Sc. and M.Sc. degrees from Mansoura University, Mansoura, Egypt, in 1994 and 1997, respectively, and the Ph.D. degree from Purdue University, West Lafayette, IN, USA, in 2004. He is a Full Professor in the School of Computing Science at Simon Fraser University (SFU), Canada, where he leads the Network Systems Lab. His 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 at 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.

Appendix A
Supplementary Materials

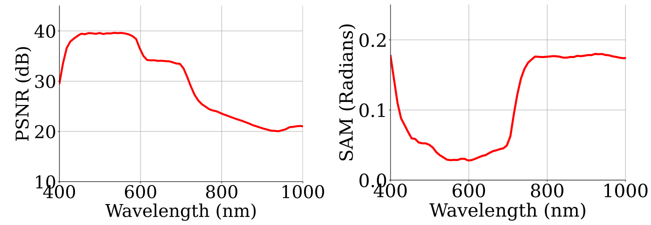
A.1 Limitations of Current Reconstruction Models

To demonstrate the limitations of current reconstruction models, we analyze the suitability of the state-of-the-art method, MST++ [21], for conducting spectral analysis to assess fruit ripeness and remaining lifetime on smartphones. MST++ is a deep neural network model and was shown to outperform all prior works [21]. We train the MST++ model on our hyperspectral fruit ripening dataset, which, as detailed in §8, has hyperspectral images of multiple fruits, and each image has 204 bands. The model takes RGB images as input and produces bands equally spaced across the visible and NIR range (400–1000 nm).

Following the guidelines for training and evaluating spectral reconstruction methods [25], we synthesize RGB images from the captured hyperspectral images using the sensitivity function of common CMOS sensors on smartphone cameras. We also assume ideal (halogen) lighting conditions. The MST++ model is trained to take RGB images as input, and it produces bands equally spaced across the 400-1000 nm range. We configured the model to reconstruct 68 bands (instead of 204) to reduce the training and inference time.

We plot the accuracy of the reconstructed bands by the MST++ model in Figure 22. The accuracy is measured using two important metrics: Peak-Signal-to-Noise-Ratio (PSNR) and Spectral Angle Mapper (SAM) [22]. The first metric measures the spatial accuracy of each reconstructed band by comparing its pixels against the corresponding ground truth band. PSNR is a quality metric, and thus, higher values are better. The second metric assesses the accuracy along the spectral dimension by measuring the angle between the spectra representing reconstructed and corresponding ground truth bands. SAM is an error metric, and thus, lower values are better. As Figure 22 shows, both the spatial and spectral accuracy quickly drop after 700 nm, which is the end of the visible light range. For example, Figure 22a shows that the PSNR of the reconstructed band at 900 nm is about 20 dB, indicating very poor quality. Similarly, Figure 22b shows that the spectral angle between the reconstructed band at 900 nm and its corresponding ground truth is 0.2 radians (11.5 degrees), which is a significant error that would compromise the accuracy of spectral signatures created from such bands.

One of the main reasons behind the poor performance of MST++ is the lack of any NIR signals in the input, which makes the reconstruction model hallucinate bands in the NIR range. We presented three possible solutions to obtain NIR signals in §5.2.



(a) Spatial accuracy (b) Spectral accuracy

Figure 22: Limitations of the state-of-the-art spectral reconstruction model [21]: It results in high errors in the NIR range.

A.2 Details of Training and Evaluating the Spectral Reconstruction Model

Training Details. For accurately training the reconstruction model, it is essential that all inputs to the model are captured in the same environment, e.g., lighting conditions, capturing distance, viewing angle, and sensor characteristics, as the ground-truth bands. Thus, similar to the standard process of training and evaluating reconstruction models [25], we use the raw reflectance data captured by the hyperspectral camera to create the input bands corresponding to ground-truth bands. Specifically, we transform the reflectance data to different inputs using the procedure described in [38]. For example, for the case of No IR Filter, we use the sensitivity function in Figure 2b to transform the reflectance data to produce images as if they were captured by the Google Nexus 5X after removing the IR filter. Recall that hyperspectral cameras capture very detailed data about the scene across many narrow bands in the spectrum. Thus, this transformation effectively *downsamples* the reflectance data and does not significantly affect the accuracy.

Similarly, for the custom color filter array, we use the sensitivity function in §5.2. For the RGB + NIR case, we use the sensitivity function in Figure 2b but truncate all signals after 700 nm, similar to what an IR filer does. Then, we randomly select one of the narrow spectral bands in the 940–980 nm as the NIR image, because most NIR cameras on phones operate in this range. Training on randomly selected NIR bands improves the robustness of our model to support diverse phones.

Performance Metrics. The details and equations of the six considered performance metrics are given below.

- *Mean Relative Absolute Error (MRAE)*: measures the absolute relative error between pixel values of the reconstructed bands $\hat{X}$ and the ground truth bands X. It is given by eq. (1).
- *Root Mean Square Error (RMSE)*: measures the second order error between pixel values of reconstructed and ground truth bands, and it is given by:

$$\sqrt{\frac{1}{HWN}\sum_{x=1}^{H}\sum_{y=1}^{W}\sum_{\lambda=1}^{N}\left|\hat{X}(x,y,\lambda)-X(x,y,\lambda)\right|^2}. \quad (5)$$

- *Spectral Angle Mapper (SAM)*: measures the similarity between the reconstructed and ground truth bands by measuring the angle between the vectors representing their spectra [22]. It is given by eq. (2).
- *Spectral Information Divergence (SID)*: models spectra as probability distributions and measures the difference between the distributions representing the reconstructed and ground truth bands [33]. It is given by eq. (3).
- *Peak Signal to Noise Ratio (PSNR)*: measures the quality of the reconstructed bands relative to the ground truth ones.

	MRAE ↓	RMSE ↓	SAM ↓	SID ↓	PSNR ↑	SSIM ↑
No IR Filter						
Pear Bosc	0.1546 ± 0.033	0.0469 ± 0.010	0.1360 ± 0.033	0.0707 ± 0.027	26.8 ± 2.0	0.9092 ± 0.029
Pear Bartlett	0.1339 ± 0.029	0.0320 ± 0.010	0.1086 ± 0.020	0.0438 ± 0.017	30.3 ± 2.5	0.9404 ± 0.023
Avocado Org	0.1299 ± 0.030	0.0217 ± 0.008	0.0762 ± 0.006	0.0124 ± 0.002	33.8 ± 3.0	0.9709 ± 0.009
Avocado Hass	0.1244 ± 0.017	0.0183 ± 0.006	0.0788 ± 0.007	0.0127 ± 0.002	35.3 ± 3.1	0.9756 ± 0.005
Average	**0.1374 ± 0.031**	**0.0310 ± 0.015**	**0.1029 ± 0.034**	**0.0378 ± 0.032**	**31.2 ± 4.5**	**0.9460 ± 0.035**
Custom Filter						
Pear Bosc	0.0590 ± 0.010	0.0101 ± 0.002	0.0574 ± 0.007	0.0059 ± 0.002	40.1 ± 1.5	0.9866 ± 0.002
Pear Bartlett	0.0638 ± 0.009	0.0096 ± 0.002	0.0599 ± 0.007	0.0065 ± 0.001	40.5 ± 1.7	0.9881 ± 0.002
Avocado Org	0.0981 ± 0.023	0.0139 ± 0.005	0.0679 ± 0.010	0.0092 ± 0.003	37.6 ± 2.6	0.9818 ± 0.007
Avocado Hass	0.0885 ± 0.012	0.0113 ± 0.003	0.0645 ± 0.008	0.0082 ± 0.002	39.2 ± 2.1	0.9856 ± 0.003
Average	**0.0761 ± 0.021**	**0.0111 ± 0.003**	**0.0620 ± 0.009**	**0.0073 ± 0.002**	**39.4 ± 2.2**	**0.9856 ± 0.004**
RGB + NIR						
Pear Bosc	0.1265 ± 0.028	0.0203 ± 0.004	0.0728 ± 0.011	0.0094 ± 0.003	34.0 ± 1.8	0.9742 ± 0.006
Pear Bartlett	0.1206 ± 0.025	0.0194 ± 0.004	0.0829 ± 0.008	0.0165 ± 0.012	34.4 ± 1.9	0.9743 ± 0.006
Avocado Org	0.1303 ± 0.021	0.0222 ± 0.006	0.0813 ± 0.011	0.0135 ± 0.004	33.3 ± 2.3	0.9723 ± 0.006
Avocado Hass	0.1102 ± 0.022	0.0204 ± 0.005	0.0852 ± 0.015	0.0155 ± 0.006	34.1 ± 2.2	0.9754 ± 0.006
Average	**0.1214 ± 0.026**	**0.0206 ± 0.005**	**0.0798 ± 0.013**	**0.0132 ± 0.007**	**34.0 ± 2.1**	**0.9742 ± 0.006**

Table 5: Performance comparison of the three versions of the spectral reconstruction model on different fruits.

It is given by:

$$10 \log_{10}(1/MSE(\hat{X}, X)), \tag{6}$$

where MSE is the average of the mean square error across all bands.

- *Structural Similarity Index Measure (SSIM)*: measures the texture similarity between the reconstructed and the ground truth bands [39]. It is given by:

$$\frac{1}{N}\sum_{\lambda=1}^{N} \mathcal{S}(\hat{X}(1:H, 1:W, \lambda)X(1:H, 1:W, \lambda)), \tag{7}$$

where $\mathcal{S}$ is the structural similarity index calculated for each band, and the procedure to compute it can be found in [40], [39].

A.3 Accuracy of Spectral Analysis

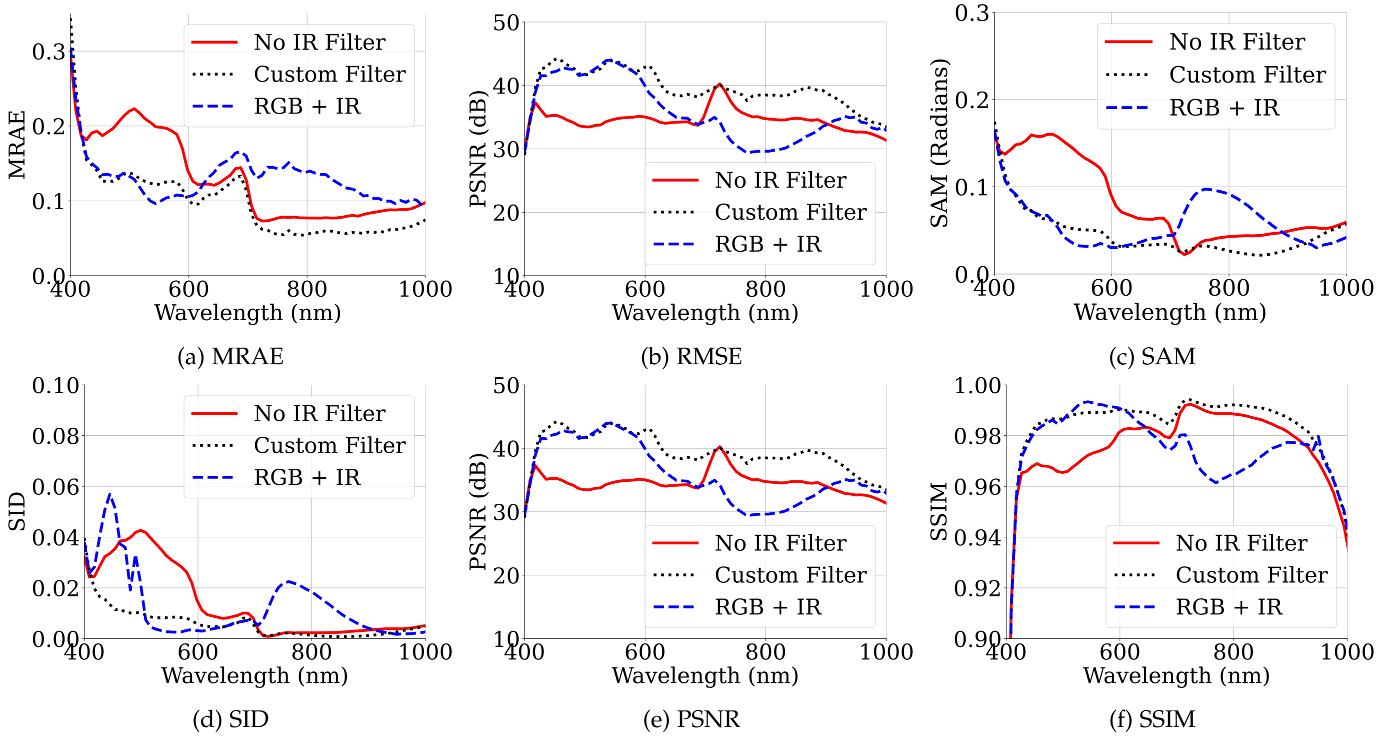
Detailed Results and Comparisons. We summarize the performance of the three versions of the reconstruction model across different fruits and all performance metrics in Table 5. Each cell shows the mean and standard deviation for the corresponding case. We note that MRAE, RMSE, SAM, and SID are error metrics. Thus, lower values are better, which is indicated by ↓ in the table. On the other hand, PSNR and SSIM represent quality metrics and higher values for them are better, which is indicated by ↑.

Multiple observations can be made on Table 5. First, all three versions of the reconstruction model produce good reconstructed bands. Specifically, all four error metrics are close to zero. For example, the average SAM value is less than 0.103 radians (5.9 degrees), and the average SID is less than 0.038. SAM and SID are particularly important for hyperspectral imaging applications since they measure the similarity between the reconstructed and ground truth bands across the spectral dimension. Similarly, the PSNR and SSIM quality metrics are fairly high. The average PSNR is more than 31 dB for all three versions of the reconstruction model, and the average SSIM approaches 1.0. Furthermore, the standard deviation of all metrics is small, indicating consistent performance.

The second observation on Table 5 is that the reconstruction model with the custom color filter array results in the highest accuracy across all metrics. This is expected as the filter is purposely designed to optimize the quality of the captured RGB and NIR bands, considering the sensor sensitivity and processing pipeline of the camera. The third observation is that the reconstruction model with RGB + NIR inputs produces better average accuracy than the model with No IR Filter. This is pleasantly surprising, considering that it does not require changing the smartphone camera.

To further analyze the accuracy of the reconstructed bands and shed some insights on the relative performance of the three versions of the reconstruction model, we plot the six performance metrics across all individual wavelength bands in Figure 23. We note that the y-axis of the error metrics in sub-figures (a)–(c) is focused on a small range since the errors are very small. This is done to demonstrate the differences among the various cases and across different bands.

Recall that the RGB + NIR case uses two separate cameras. The RGB camera still uses an IR filter and thus retains the quality of RGB images. This yields a higher reconstruction accuracy in the visible light range compared to the No IR Filter case where the quality of the RGB images is damaged because of the interference with the IR signals. This is shown in the left parts (between 400 and 700 nm) in the sub-figures of Figure 23. Further, the additional NIR image in this case, 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. However, the



(a) MRAE (b) RMSE (c) SAM

(d) SID (e) PSNR (f) SSIM

Figure 23: Detailed analysis of the spectral reconstruction model when using three possible inputs: (i) RGB images captured with No IR Filter, (ii) Images captured with the Custom Filter, and (iii) RGB + NIR images captured by unmodified phones.

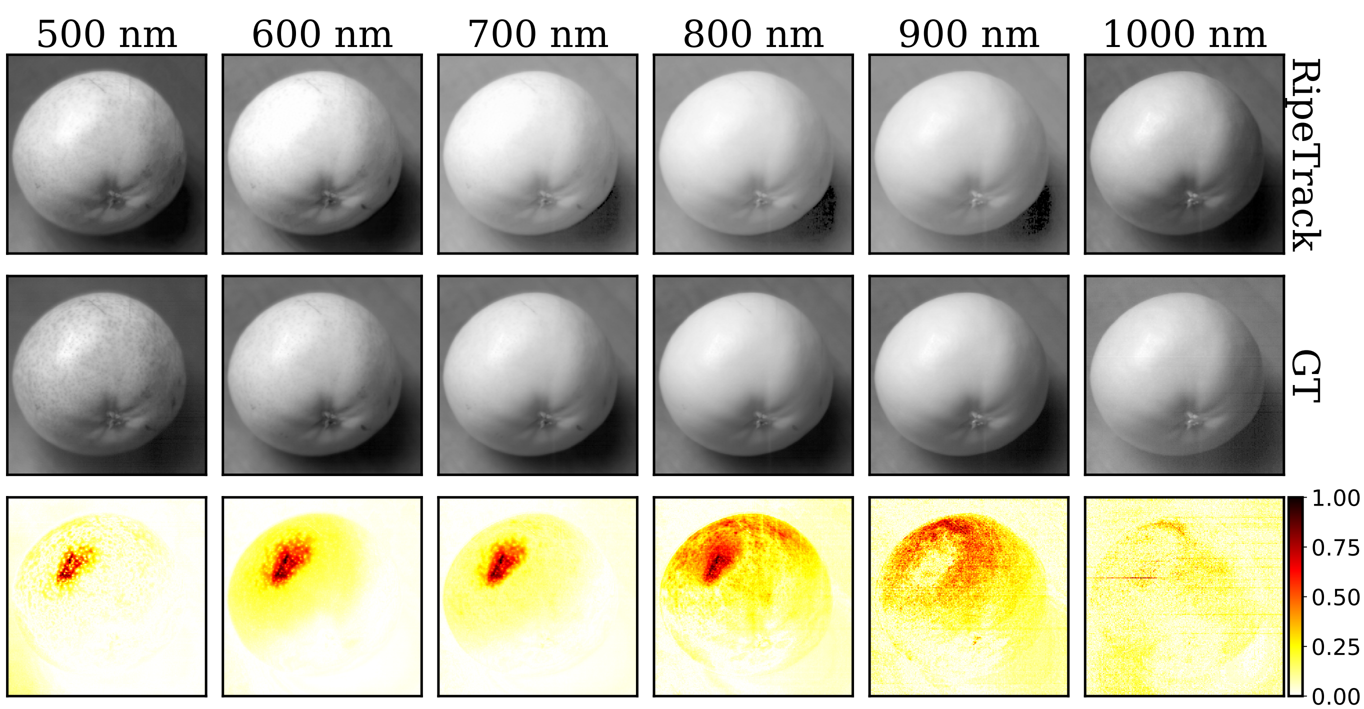


Figure 24: Visual comparison of the reconstructed bands and the ground truth (GT) ones; the bottom row shows the absolute errors between them as heat maps. Results shown for a sample Pear Bartlett.

RGB + NIR case has relatively higher errors and lower quality than the No IR Filter case in the 750–850 nm range. This is because the transition from the visible range to the NIR range occurs around the 700 nm band, where all RGB signals are truncated. Thus, the reconstruction model has less information to build on in the 750–850 nm range, leading to higher errors.

In addition, the custom color filter array provides lower errors and higher quality across most bands as shown in Figure 23. This is because, as illustrated in §5.2, this filter provides a *wider* NIR band around 850 nm and does not truncate the RGB signals at 700 nm, providing more information to the reconstruction model throughout the entire 400–1000 nm range.

Visual Samples of the Reconstructed Bands. We provide additional samples demonstrating the quality of the reconstructed bands in Figure 24. The figure also shows the difference between each reconstructed band and its corresponding ground one as a heat map.

Summary. The presented spectral reconstruction model produces fairly accurate bands in all three considered cases. The custom color filter array provides the highest accuracy, but it requires significant changes to the camera sensor. Removing the IR filter results in good reconstruction, but it damages the RGB images. Using RGB and NIR images offers a practical solution, providing high reconstruction accuracy without requiring any hardware modifications or damaging the RGB images.