



# FruitPhone: Detecting Sugar Content in Fruits Using Unmodified Smartphones with Spectral Imaging

HAIYAN HU, Hong Kong University of Science and Technology, China

YINAN ZHU, Hong Kong University of Science and Technology, China

SHANWEN CHEN, Hong Kong University of Science and Technology, China

QIANYI HUANG, Sun Yat-sen University, China

QIAN ZHANG\*, Hong Kong University of Science and Technology, China

Developing an accessible fruit sugar content detection system for average users, particularly diabetic patients, is essential. Current fruit sugar detection solutions often fail to balance ease of use with high accuracy. In this study, we introduce FruitPhone, an innovative smartphone-based spectral imaging system that provides quantitative assessments of fruit sugar content. To enable the smartphone to capture the fine-grained spectral features necessary for accurate sugar content analysis, FruitPhone utilizes the smartphone screen to simulate multiple monochromatic light sources. This approach allows for the collection of spectral data that extends beyond the standard three-channel RGB format. To address the discrepancies between screen-generated and true monochromatic light, we propose a two-stage spectral reconstruction algorithm that translates the smartphone's pseudo-spectral data into real multi-spectral images for further analysis. Additionally, we tackle environmental variability by employing a black screen light for reference data collection, which helps mitigate ambient influences and enhance consistency. We also introduce a detrending algorithm for high-dimensional spectral images to correct spatial inconsistencies while preserving spectral trends. We evaluate FruitPhone using 335 spectral images from 37 different fruit types, achieving a normalized mean-absolute error of  $8.48\% \cdot Bx$ , which is comparable to that of a laboratory spectral system and demonstrates a 19.98% reduction in reconstruction errors, along with a 45.58%  $R^2$  improvement over baseline schemes.

CCS Concepts: • **Human-centered computing** → **Nutrient estimation; Mobile devices**; • **Computing methodologies** → **Machine learning**.

Additional Key Words and Phrases: smartphone, fruit sugar content, spectral imaging, spectra reconstruction

## ACM Reference Format:

Haiyan Hu, Yanan Zhu, Shanwen Chen, Qianyi Huang, and Qian Zhang. 2025. FruitPhone: Detecting Sugar Content in Fruits Using Unmodified Smartphones with Spectral Imaging. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 9, 3, Article 86 (September 2025), 29 pages. <https://doi.org/10.1145/3749470>

\*Corresponding author.

Authors' Contact Information: [Haiyan Hu](#), Hong Kong University of Science and Technology, Hong Kong, China, [hhuap@connect.ust.hk](mailto:hhuap@connect.ust.hk); [Yinan Zhu](#), Hong Kong University of Science and Technology, Hong Kong, China, [yzhudf@cse.ust.hk](mailto:yzhudf@cse.ust.hk); [Shanwen Chen](#), Hong Kong University of Science and Technology, Hong Kong, China, [schenfh@connect.ust.hk](mailto:schenfh@connect.ust.hk); [Qianyi Huang](#), Sun Yat-sen University, Guang Zhou, China, [huangqy89@mail.sysu.edu.cn](mailto:huangqy89@mail.sysu.edu.cn); [Qian Zhang](#), Hong Kong University of Science and Technology, Hong Kong, China, [qianzh@cse.ust.hk](mailto:qianzh@cse.ust.hk).

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM 2474-9567/2025/9-ART86

<https://doi.org/10.1145/3749470>

## 1 INTRODUCTION

Brix, a measure of sugar content in fruits, directly influences taste and quality. The need for an easy-access fruit brix detection tool is highlighted for average users during purchasing. According to a study by the United States Department of Agriculture (USDA), over 60% of consumers prioritize sweetness when selecting fruits, making brix measurement vital for informed choices [50]. Reports indicate that approximately 1.3 billion tons of food are wasted annually due to uninformed purchasing decisions [18]. Additionally, developing an easy-access fruit brix detection solution may also help reduce healthcare costs related to diet-related health issues, such as diabetes [47]. Research indicates that 70% consumers express concern over food quality but often lack the knowledge to accurately assess sugar content in fruits [27], leading to unintentional purchases of higher-sugar options that may not fit their dietary needs. Therefore, an easy-to-use tool can empower users to check fruit quality while shopping, which can not only enhance the shopping experience but also promote healthier eating habits and reduce food waste.

However, existing solutions for fruit brix detection can not meet the requirements of easy accessibility and high accuracy simultaneously. Traditional methods like hydrometers, refractometers, and electronic tongues provide precise sugar content measurements but require sample destruction for liquid extraction, making them impractical to be used for average consumers during shopping [40]. In contrast, non-destructive methods such as radio frequency and hyperspectral imaging analyze the radio or optical absorption characteristics of fruit to determine sugar levels without damaging the sample. However, the high cost of existing non-destructive solutions is a significant barrier to their widespread adoption for average users, which can cost \$300 to \$10,000 [16, 54]. To further reduce cost and enhance the accessibility, some recent studies have tried to achieve smartphone-based spectrometers [55, 64]. However, these solutions either require additional hardware support [22, 64] or have limited spectral resolutions [55], which hinders their application in fine-grained quantitative spectral analysis.

In this work, we explore whether we can achieve quantitative fruit brix detection using only an unmodified commercial smartphone. Our key idea is to reuse the optical sources and sensors in the smartphone to build an accurate spectral imaging system. While some previous studies have attempted to develop smartphone-based spectrometers [55, 57, 58, 64, 70], they have not been effective in the context of fruit sugar estimation. Specifically, some approaches focus on reconstructing smartphone RGB images into hyperspectral images using deep learning reconstruction algorithms [55] for fruit quality detection. However, the challenge is that estimating spectral information from only RGB colors is an ill-conditioned one-to-many mapping problem, especially for metameric colors [19, 73], limiting the quality of the reconstructed spectra. The study by [64] attempts to address this issue by introducing an additional LED array, which reduces the versatility of the solution. Other research aims to use the smartphone screen to simulate multiple colors and generate a new pseudo color spectrum, which proves effective for identifying metameric materials [57, 58, 70]. However, since smartphone screens consist of RGB LEDs rather than monochromatic LEDs, the resulting color spectra can only be employed for colorimetric analysis and cannot be directly applied for quantitative spectral analysis.

To bridge this gap, we propose reconstructing the color spectra illuminated by smartphone screen light sources into a real hyperspectral image. The advantage of using color spectra instead of RGB images for reconstruction is that they contain many more monochromatic channels. Previous studies have demonstrated that reconstruction can be more reliable with a greater number of input channels [19, 73]. Therefore, by modulating and time-multiplexing the smartphone screen, we can capture images across dozens of different channels, thereby enhancing the quality and reliability of spectral reconstruction.

However, putting this idea into practice is not straightforward due to several significant challenges:

- **Limitations of Screen-Simulated Light.** A smartphone screen consists of structured red, green, and blue LEDs, which means that the simulated monochromatic light is actually a combination of these colors in varying proportions. While this approach can provide more detailed spectral information than

standard three-channel RGB images, the functioning of each channel is fundamentally different from that of traditional multi-spectral images. This difference makes existing reconstruction algorithms ineffective, as they rely on a clear and consistent interpretation of each channel.

- **Impact of Uncontrolled Environments on Data Quality.** Imagine trying to take a picture of a fruit under a bright light, with shadows dancing around it, or at different distances where the light might not hit evenly. Factors like varying ambient light conditions and the distance between the smartphone camera and the fruit can create instability and uneven light distribution in the collected data, leading to challenges in achieving accurate and reliable results.

To address the challenges outlined above, we propose FruitPhone, a smartphone-based spectral imaging system that can accurately quantify the sugar content of various fruits. **For the first challenge**, we recognize that, according to CIE-RGB color space mapping [65], screen-simulated monochromatic light can be viewed as a mapping from a real monochromatic sensor to RGB LEDs. Therefore, we can first establish an inverse mapping to translate the collected data from pseudo-spectral space to real spectral space while maintaining the same number of channels. This approach allows us to increase the number of input channels available for spectral reconstruction, thereby enhancing reconstruction reliability. Specifically, we introduce a two-stage spectral reconstruction algorithm: the first stage translates the collected pseudo-spectral images into real multi-spectral images using a three-branch transformer module, and the second stage employs a state-of-the-art spectral reconstruction model to derive full spectra from the translated features. **For the second challenge**, to mitigate disturbances from surrounding environments, such as varying light conditions, we incorporate reference data captured with a black screen light. Since this reference data experiences the same environmental influences as the other data, we can effectively eliminate disturbances by subtracting the reference. Additionally, to address light unevenness caused by variations in light sources and collection distances, we propose a detrending algorithm for high-dimensional spectral images that reduces spatial unevenness while preserving spectral trends.

We implement FruitPhone on a commercial smartphone, Google Pixel 4 XL. Given the informative absorption bands of sugar in the near-infrared (NIR) range, we incorporate NIR images during the second step of our spectral reconstruction process. We evaluate the effectiveness of FruitPhone using 335 spectral images collected from 37 different types of fruits. Our system achieves a normalized mean-absolute error (NMAE) of 8.48% °Bx, which is only 1.3% °Bx higher than that of an expensive laboratory hyperspectral spectrometer (priced over \$10,000). Given the accepted quality estimation error in the food industry is 10% [1], the error from FruitPhone has a negligible impact on taste and is imperceptible to consumers. Compared to the state-of-the-art smartphone-based spectral system baseline [55], our solution reduces reconstruction errors by 19.98% and increases the  $R^2$  value for brix regression by 45.58%. Additionally, FruitPhone demonstrates remarkable robustness across various setups, including diverse fruit placements, user configurations, smartphone models, and uncontrolled real-world scenarios.

We summarize the contributions as follows:

- We develop the first system that enables accurate, quantitative assessment of fruit sugar content using only a smartphone, making this technology more accessible than ever.
- By utilizing the smartphone screen to simulate multiple monochromatic light sources, we address the limitation of smartphones providing only three-channel input. This approach, combined with our innovative two-stage spectral reconstruction algorithm, significantly enhances spectral reconstruction quality, enabling quantitative detection using only smartphone-integrated components.
- We implement FruitPhone on a commercial smartphone and evaluate its effectiveness across various fruit types and environmental conditions, demonstrating its reliability in real-world scenarios.

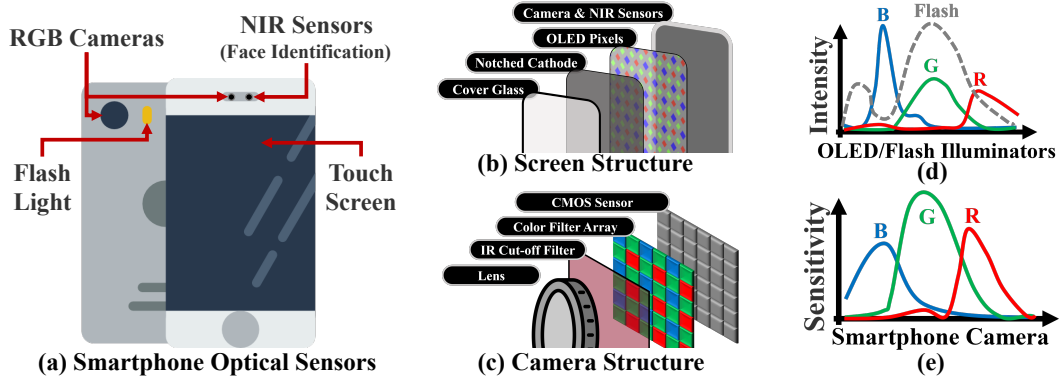


Fig. 1. Built-in optical sources and sensors of a commercial smartphone (a). Physical structure of the smartphone screen (b) and camera (c), which both contain no more than three wavelength channels (d-e).

## 2 BACKGROUND AND FEASIBILITY STUDY

### 2.1 Spectral Imaging for Fruit Brix Detection

Spectral imaging technology is a powerful tool to assess fruit quality such as ripeness, sugar content, and internal defects in a nondestructive way [51, 71]. It analyzes the unique absorption and reflectance characteristics of light with different wavelengths as they interact with fruit tissues to infer the concentration of the components.

According to Beer's Law [61],  $A = \epsilon l \cdot C$ , the fruit absorbance spectrum  $A$  is linearly related to the concentration of the components  $C$ . The  $\epsilon$  and  $l$  are the molar attenuation coefficient and the optical path length of the light, which are effected by the physical structure of the sample. The spectrometer is used to obtain sample's absorbance or reflectance spectrum, whose output is formulated as,

$$I_\lambda = I(x, y, \lambda) = E(\lambda) \cdot R(x, y, \lambda) \cdot S(\lambda) = E \cdot R \cdot S, \quad (1)$$

where  $R$  is the samples' reflectance spectrum that can be transferred into absorbance with  $A = \log(1/R)$ ,  $x$ ,  $y$  and  $w$  are the width, height and wavelength channel respectively.  $E$  and  $S$  are the light source radiation and detector sensitivity of the spectrometer hardware. Therefore, we can build a relationship between the spectrometer output  $I_\lambda$  and the components' concentration  $C$ , i.e.,  $C = \phi(I_\lambda; E, S, \epsilon l)$ . Thus, we can identify and quantify fruit's sugar content by analyzing its spectral absorption.

In the context of fruit brix detection, some wavelengths in the near-infrared (NIR) and visible range have been identified as critical for measuring sugar content [29, 29, 45]. For example, 740 to 800 nm correspond to vibrational overtones and combinations of molecular bonds present in sugars (e.g., C-H and O-H bonds) [63], 680 nm and 720 nm in the visible range can also provide insights into chlorophyll content [10]. By analyzing the absorption spectra, spectroscopy can deliver accurate assessments between varying levels of sweetness in fruits while maintaining their integrity. However, the high cost of existing spectral imaging solutions is a significant barrier to their widespread adoption for average users. A commercial spectrometer for fruit sugar content can cost \$300 to \$10,000 [16, 54], making it inaccessible for smaller operations and individual users. In contrast, a smartphone-based fruit sugar content detection system offers cost-effectiveness, portability, and user-friendliness compared to existing solutions.

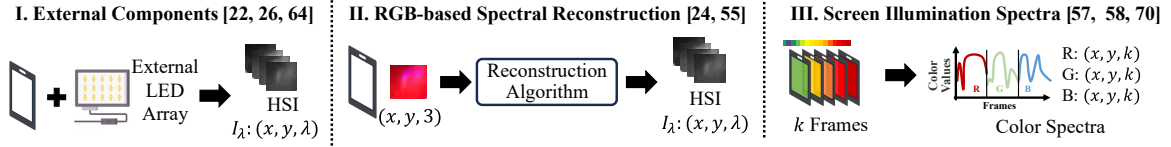


Fig. 2. Three types of existing smartphone-based spectrometers, relying on different approaches to extend the number of wavelength channels in the smartphone's output: (I) external components, (II) RGB-based spectral reconstruction, and (III) screen illumination spectra.

## 2.2 Smartphone Optical Sensors

Smartphones are equipped with various optical components that enhance their functionality, as shown in Figure 1(a). These optical sensors can be taken as light sources or detectors for a spectral imaging system.

**Light Sources in Smartphones.** Smartphones are equipped with two primary light sources: the flash light and the display screen. The flash light, typically an white LED, provides a brief, high-intensity burst of illumination. The display screen, often an LCD or OLED, consists of several layers including color filters (as shown in Figure 1(b)), allowing it to emit a range of colors and intensities. This versatility enables the display to serve as a continuous light source, adjustable for various imaging scenarios. Figure 1(d) is an example of the illumination spectrum for the two light sources. Both light sources can be repurposed for spectral imaging; the flash light is effective for capturing quick, bright images, while the display allows for controlled illumination across different spectral bands. However, due to the lack of splitter, they are limited to provide fine-grained and continuous monochromatic light.

**Cameras in Smartphones.** Cameras in smartphones typically include front-facing and rear-facing RGB cameras. Some recent phones, like Apple iPhoneX and Google Pixel4, involving a near-infrared camera for face identification. Figure 1(c) shows the structure of a commercial smartphone, which consists of a lens system that focuses light onto a Complementary Metal-Oxide-Semiconductor (CMOS) image sensor covered with a color filter array (CFA), allowing it to capture red, green, and blue components of visible light. The spectral sensitivity of the CMOS image sensor is shown in Figure 1(e). This camera is optimized for full-color imaging but is limited to the visible spectrum. In contrast, the NIR camera is designed to capture light in the near-infrared spectrum, typically from 700 nm to 1100 nm. While both cameras can serve as light detectors in a spectral imaging system, the RGB camera is effective for visible spectrum analysis, and the NIR camera extends capabilities into the near-infrared range.

## 2.3 Smartphone-based Spectrometers

Although equipped with extensive optical sensors, smartphones still lack an effective way to generate fine-grained and continuous wavelength spectra. In comparison to the hyperspectral images in Eq 1, the RGB images obtained by a smartphone contain only three wavelength channels, represented as  $I_c(x, y, 3) = [I_r(x, y), I_g(x, y), I_b(x, y)]$ . Each channel can be formulated as:

$$I_c(x, y) = \int_{\lambda} E(\lambda) \cdot R(x, y, \lambda) \cdot S_c(\lambda), \quad (2)$$

where  $S_c$  is the smartphone's CMOS sensitivity for the corresponding color. Therefore, previous works have attempted to extend the number of wavelength channels in the smartphone's output. As illustrated in Figure 2, there are three main approaches.

**External Components.** One straightforward idea is to add external illumination or detection sensors to the smartphone [22, 26, 64]. By adding an LED array with different wavelengths ( $E_{\lambda_1}, E_{\lambda_2}, \dots, E_{\lambda}$ ) and illuminating them through time division multiplexing, we can capture a series of images representing various wavelengths, *i.e.*,

$I_{\lambda_1}, I_{\lambda_2}, \dots, I_{\lambda}$ . However, the spectral image obtained by this method has the same number of bands as the external LED array, making it challenging to achieve hundreds of bands like those found in commercial hyperspectral cameras. Although [64] addresses this issue by employing spectral reconstruction algorithms, the necessity for additional hardware still limits its versatility for average users.

**RGB-based Spectral Reconstruction.** Another idea is to utilize machine learning algorithms to reconstruct hyperspectral images from RGB images captured by the smartphone [24, 55]. The rationale is to build a mapping matrix from the smartphone's hardware parameters to the hyperspectral system's, *i.e.*,  $I_{\lambda} = M \times I_c$ . However, the challenge is that estimating spectral information from only RGB colors is an ill-conditioned one-to-many mapping problem, especially for metameric colors [19, 73], limiting the quality of the reconstructed spectra. [55] utilizes the NIR sensor on mobile phones, expanding the reconstruction channels to 4 (RGB+NIR) and successfully restoring 68 channels of hyperspectral data for fruit quality inspection. However, the ill-posed nature of the problem still limits reconstruction performance, making it suitable only for basic classification tasks rather than for precise measurements, such as determining sugar content.

**Screen Illumination Spectra.** Some previous work also tried to generate more than three colors using the smartphone screen [57, 58, 70]. By mimicking  $k$  monochromatic colors using the smartphone screen and playing them in sequence, we can obtain a color spectrum,  $I_s(3, x, y, k)$ . It contains multiple frames, each of which is related to one monochromatic color. This color spectrum is validated to have good performance in identifying various substances with metameric colors, showing promise in qualitative tasks, such as material classification [70] and food adulteration detection [58]. However, the color spectrum differs physically from real absorption spectra, making it unsuitable for direct component quantification.

## 2.4 CIE-RGB Color System

FruitPhone proposes to utilize the screen sensors to simulate monochromatic light at various wavelengths for the spectral imaging system. One of the key steps is to translate the wavelength of monochromatic light into R,G,B values. To achieve this, we need to leverage the CIE-RGB color system, a color representation model developed by the International Commission on Illumination (CIE). According to the CIE-RGB color system, each color can be represented as a combination of these three primary colors, with values typically ranging from 0 to 255 in digital displays. Specifically, to translate the wavelength  $\lambda$  into  $(R, G, B)$  values, we need the following three steps. (1) Translate wavelengths to CIE color coordinates  $(x, y)$  based on the CIE 1931 color space [65]. And define the illumination value  $Y = 1$ . (2) Convert the chromaticity coordinates  $(x, y)$  to RGB using the following formulas:

$$R = Y \cdot \frac{x}{y}, G = Y \cdot \frac{1 - x - y}{y}, B = Y \cdot \frac{(1 - y)}{y}. \quad (3)$$

(3) Map CIE-RGB values to the device's RGB color space according to the display's color gamut and characteristics. Figure 3 shows CIE-RGB color space chromaticity diagram and the translation of monochromatic light spectra to smartphone color spectra. Although the color spectrum is pseudo, it provides more spectral information than the original RGB images, showing differentiation for metameric colors. On the other hand, the increased number of monochromatic wavelengths alleviates the ill-conditioned one-to-many mapping problem in RGB-based spectral reconstruction [73]. However, reconstructing hyperspectral images from smartphone color spectra is still challenging, which is the target of this work.

## 3 FEASIBILITY STUDY

We conduct a preliminary experiment in spectral space to assess the feasibility of FruitPhone. Our primary objective was to validate two hypotheses: (1) The pseudo-spectrum generated by monochromatic light simulated through the phone screen contains more spectral information than the RGB images. (2) An increase in the number of input spectral channels can improve the accuracy of spectral reconstruction.

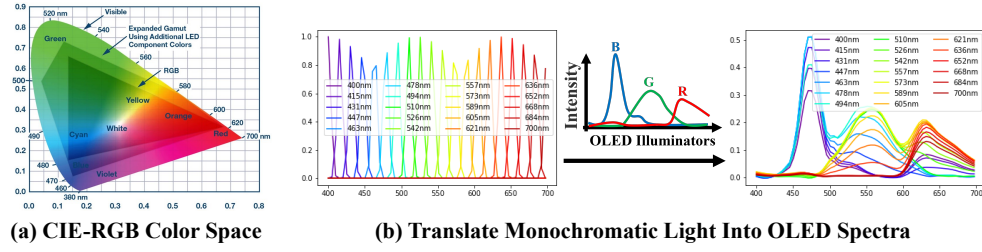


Fig. 3. Illustration of CIE-RGB color space and its utilization of translating monochromatic light into smartphone OLED monochromatic spectra.

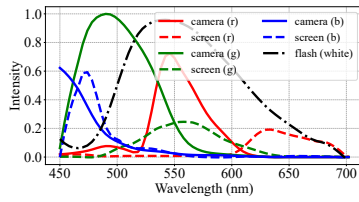


Fig. 4. Optical sensor parameters we used for feasibility study.

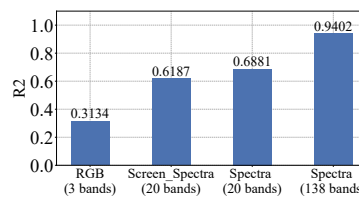


Fig. 5. Performance of fruit brix regression using various input spectra.

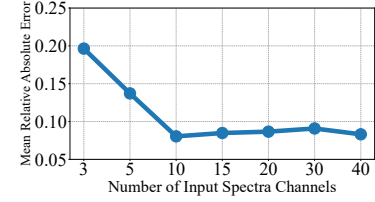


Fig. 6. Spectral reconstruction errors of using various number of spectra channels.

To validate these hypotheses, we collect 4,000 one-dimensional absorption spectra ranging from 400 to 1000 nm from 200 fruit samples across 10 different types. We utilized the camera and OLED screen parameters published by manufacturers to simulate various spectra [41]. Figure 4 illustrates the parameters used for this simulation. Specifically, the RGB spectra were calculated by multiplying the fruit spectra  $R$  with the smartphone flash emission  $E_{flash}$  and the RGB camera response  $S_p$ , resulting in a shape of  $(n, 3)$ , where  $n$  is the number of samples. We then uniformly selected 20 wavelengths within the range of 400 to 700 nm and generated screen emission spectra through CIE-RGB color space translation. The screen spectra were calculated by multiplying the screen emission  $E_{screen}$  with the fruit spectra and the camera parameters, yielding a shape of  $(n, 20, 3)$ . The multi-spectra utilized a light source at the same wavelengths as the screen spectra, resulting in a shape of  $(n, 20)$ . We employed standard normal variate (SNV) normalization to standardize the spectra to the same intensity scale before inputting them into a partial least squares (PLS) regression model to predict sugar content.

Figure 5 presents the results of sugar content regression using various spectral types. The RGB data, which contains only three channels, exhibited the poorest performance, while the full-band spectra with 138 channels achieved an  $R^2$  value of 0.94. The screen pseudo-spectra and real spectra demonstrated comparable performances with  $R^2$  values of approximately 0.62 and 0.68, respectively. These results validate that the simulated screen spectra contain more informative spectral information than the RGB data, suggesting a potential for improved performance in spectral reconstruction. Furthermore, we employed a linear regression (LR) model, which is widely used as baseline for one-dimensional spectral reconstruction [25], to reconstruct the 138-band spectra using down-sampled data with varying channel numbers. Figure 6 illustrates the reconstruction error for various number of input spectra channels. We can clearly observe that when the input spectra contains few channels, *i.e.*, less than 5, the reconstruction errors are much larger. Furthermore, a clear decreasing trend in reconstruction

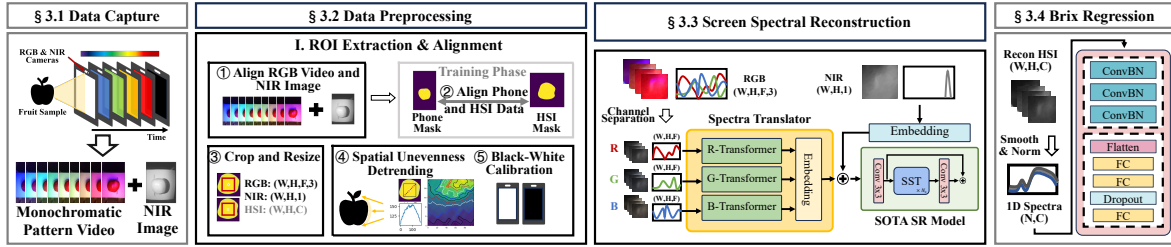


Fig. 7. Illustration of the architecture of FruitPhone, which includes four modules: data capture module, data preprocessing module, screen spectral reconstruction module, and brix regression module.

error is observed as more channels are utilized, underscoring the importance of having sufficient channels for accurate and reliable spectral reconstruction.

## 4 SYSTEM DESIGN

This section elaborates on the system design of FruitPhone. Figure 7 illustrates an overview of FruitPhone's system design. Specifically, the data capture module (§ 4.1) introduces the smartphone app that scans the fruit sample with varying monochromatic light while capturing a monochromatic pattern video and a near-infrared image. Then, we feed the captured RGB video and NIR image into the data processing module (§ 4.2) to extract the region of interest (ROI) and align it in the spatial dimension. Additionally, we reduce the noise caused by spatial unevenness and surrounding interference from the data. Subsequently, the cropped and aligned RGB video and NIR image are input to the screen spectra reconstruction module (§ 4.3) to reconstruct full-band hyperspectral images using a two-stage transformer-based reconstruction model. Finally, after smoothing and normalization, the fruit spectra are used by a convolutional neural network (CNN)-based regression model to accurately predict the sugar content of the fruit sample (§ 4.4).

### 4.1 Data Collection

Compared with the traditional spectral reconstruction work based on mobile phones [23, 55], which only uses the data from the three RGB channels for reconstruction, FruitPhone utilizes the phone screen to simulate monochromatic light of different wavelengths to obtain more wavelength channels. Specifically, we uniformly select several different wavelengths from the visible range (i.e., 400-700 nm) and calculate their RGB values according to the CIE-RGB Color system. It is worth noting that FruitPhone can change the mode of the monochromatic light on the screen to achieve better results according to the spectral absorption characteristics of the object being analyzed. For example, if carotenoids exhibit absorption characteristics at 460-520 nm, the number of bands near 460-520 nm in the monochromatic light mode can be increased. In addition, the value of  $K$  is linked to the acquisition time and the resolution of the phone. A larger  $K$  value facilitates reconstruction but means a longer acquisition time. In this paper, we define  $K = 45$ , and the acquisition time is about 2 seconds, which is easy for users to stay still while can obtaining enough monochromatic bands. Additionally, we added a white light and a black light before and after the  $K$  monochromatic pattern, respectively, as calibration data to eliminate environmental effects. We developed an app that uses the front camera of the phone to record while the screen emits monochromatic light, and then activates the front NIR camera to take an NIR picture after the recording is completed. We then extract key data frames from the recorded video, each of which captures a fruit sample illuminated by a different monochromatic light.

## 4.2 Data Preprocessing

**4.2.1 ROI Extraction and Data Alignment.** The first step after data capturing is to achieve pixel-to-pixel alignment between the smartphone RGB video, NIR image, and the HSI image, which is necessary for spectral reconstruction. Furthermore, to reduce processing time, we focus on the most informative area and crop the central areas of fruit samples for further analysis. Specifically, the pipeline contains the following steps.

**Align RGB Video and NIR Image.** Since both the field-of-view and shooting angle of the smartphone RGB and NIR cameras are different, it's necessary to align the RGB video data and NIR image before extracting ROI for further analysis. Fortunately, because the parameters of the two cameras are fixed, we only need to calculate a mapping relationship ( $\mathcal{M}$ ) with a small number of samples, which can be applied to other data. In this case, we first use several sets of images to learn the alignment matrix ( $\mathcal{M}$ ), including rotation, translation, padding, and scaling, and apply  $\mathcal{M}$  to all data. We then leverage the state-of-the-art segmentation model, Segment Anything Model (SAM) [30], to extract ROI regions from the aligned data. When extracting ROI from RGB data, we find two types of identification errors: incomplete extraction, and additional shaded areas. We avoid the former by setting a size threshold for the extraction area. For the second problem, we observe that the contrast between shadows and objects in NIR images is higher than in RGB data. Therefore, we extract ROI from aligned RGB data and NIR data respectively, then merge the extracted ROI masks, and extract the common part as the final ROI region. Using this method, the ROI region of the aligned data has an average Intersection over Union (IOU) of more than 0.97.

**Align Smartphone and HSI Data.** In addition, pixel-to-pixel aligned smartphone and HSI data are required to train the spectral reconstruction model. The alignment pipeline contains three steps. (1) Extract ROI and central circle. We first employ the SAM to accurately identify and isolate the ROI, simultaneously generating a binary mask to highlight the selected area. (2) Central Alignment. We subsequently calculate the maximum inner circle of the ROI to achieve image size adjustment. We adjust the image size according to the inner circle radius, and use linear interpolation to obtain the proper size for further analysis. (3) Rotation and Alignment. We then perform rotation to align the angle of the two ROI areas. The rotation angle is decided with the largest IOU during the rotation process.

**Cropping and Resize.** Following this, to avoid potential shadows during shooting, we take 80% of the internal area after cropping as our final area to extract. Subsequently, we adjust the dimensions of the cropped smartphone and HSI regions to meet the necessary size requirements for further analysis. Specifically, the extracted RGB size is  $(W, H, F, 3)$ , the NIR image is  $(W, H, 1)$ , and the HSI image is  $(W, H, C)$ , where  $W, H$  is the width and height of the image,  $F$  is the number of frames of the RGB video, and  $C$  is the number of spectral channels for the HSI data.

Based on the above procedure, we can ensure the processed smartphone data and HSI images are well paired regardless of the two cameras' shooting locations, angles or resolutions.

**4.2.2 Data Noise Cancellation.** The principle of spectral reconstruction is to learn the mapping relationship between mobile phone data and hyperspectral data in the spectral dimension. Therefore, removing the unstable noise from the input data is an important prerequisite for learning stable reconstruction mappings. By analyzing the data, we find that there are two main sources of noise in the data: uneven distribution of the spatial light field and environmental interference. We eliminate the two types of noise by specific approaches.

**Detrend Spatial Unevenness.** As shown in Figure 8, due to the curvature of the fruit sample, the distance and angle of the light emitted by the mobile phone screen to the sample are different, resulting in uneven intensity in the spatial dimension. In particular, the brightness of the middle is greater, and the brightness of the surrounding is smaller. For example, Figure 9 shows one frame of an apple's RGB video, while the blue line in the left plot of Figure 10 shows the brightness of different pixels on the diagonal, where the horizontal axis represents the pixels from the top left to the bottom right, and the vertical axis represents the mean value of the three channels of the pixel. We can clearly see that the intensity of the image is the strongest in the middle, and the two-way

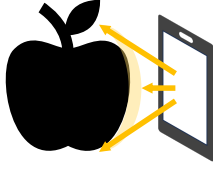


Fig. 8. The spatial unevenness of the sample.

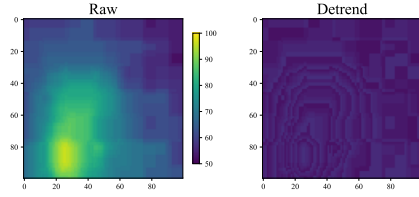


Fig. 9. The first frame of RGB video data with and without 2d-detrending process.

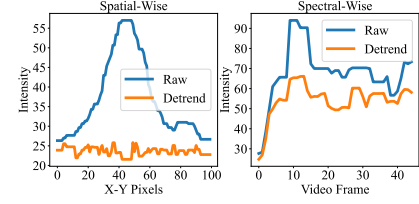


Fig. 10. The intensity spectrum of the RGB video data along the x-y pixels and frames.

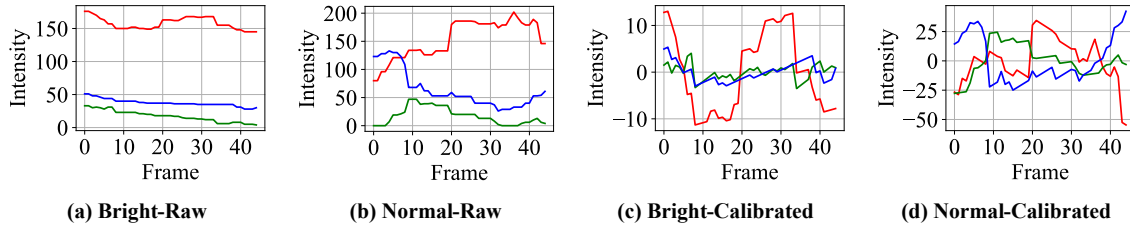


Fig. 11. The R,G,B channel intensity spectra under various light conditions before and after calibration.

weakening is reflected in the spectral dimension, which shows a huge difference in the spectrum of different regions of the same fruit sample.

To this end, we design a detrending scheme for two-dimensional data, which aims to eliminate the intensity differences in the spatial dimensions of the samples, while preserving the intensity correlation in the spectral dimensions. The core idea is to convert the detrending problem of two-dimensional data into a one-dimensional detrending problem. Specifically, the detrending process involves automatically clustering pixels into intensity-based gradient levels. In this study, we set the number of gradient levels to 10, optimizing this choice to balance computational efficiency with effective noise reduction. This selection is empirically validated, as using fewer levels resulted in insufficient correction of intensity drifts, while increasing the number of levels introduced unnecessary computational costs without providing additional denoising benefits. Following this, we apply one-dimensional detrending to the average intensity array to obtain the intercept  $b$  and slope  $k$ . We then calculate the trending map by multiplying the granularity by the slope  $k$  and subtract this trending map from the raw images. For example, for an image with an intensity range of  $[I_{min}, I_{max}]$ , we divide this range into 10 equal-width intervals. Each pixel is assigned to its corresponding interval based on its intensity value. We then compute the mean intensity for each interval and perform one-dimensional detrending on the average intensity array. Finally, we apply the detrending to each pixel using the formula:  $I_{detrend}(x, y) = I_{raw}(x, y) - (m_i \cdot I_{raw}(x, y) + b_i)$ , where  $i$  denotes the interval to which pixel  $(x, y)$  belongs. Figures 9 and 10 show the spatial and spectral dimensions of the original and de-trended data, from which we can see that the processed data narrows the differences in spatial dimensions while preserving the change trend of spectral dimensions.

**Reduce Surrounding Interference.** Since FruitPhone is supposed to be used under various environments, interference in the environment, such as ambient light or user jitter, may affect the quality of the collected data. Figure 11(a-b) shows the intensity spectrum of three channels of RGB video data R, G, and B of the same sample collected under two different ambient lights, namely bright (400 lx) and normal (120 lx). From the figure, we can clearly see that the strong ambient light will drown out the trend of the data with the change of the screen light mode. Therefore, these noises need to be removed from the data to obtain the correct monochromatic light

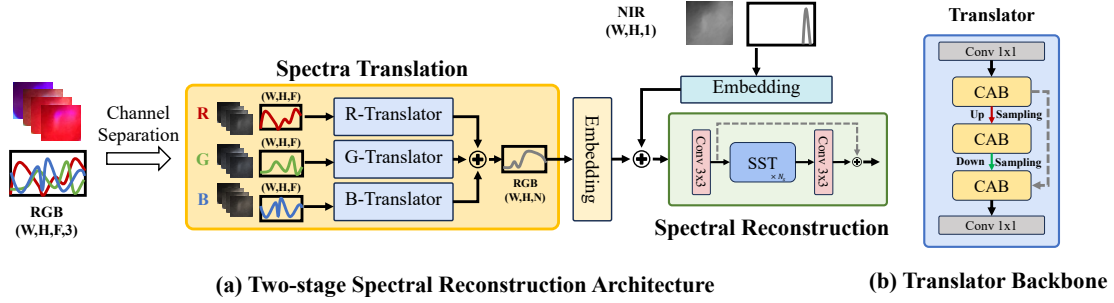


Fig. 12. The architecture of our proposed two-stage spectral reconstruction model, which separates the screen spectral reconstruction problem into two tasks: spectral translation and spectral reconstruction.

pattern. To this end, FruitPhone inserts reference video frames illuminated by white and black light sources at the beginning and end of the captured RGB video. According to the display principle of OLED, the white light source represents R, G, and B sensor brightness is 1, and the black light source represents R, G, and B sensor brightness is 0. Therefore, the two reference frames collected can be used as reference values for black and white calibration. We first subtract the black reference frame from the raw data to eliminate the basic intensity from the background, then divide by the trend of brightness change from the white reference frame to the black reference frame to compensate for the reduced intensity of the phone screen light source. Figure 11(c-d) illustrates the RGB video intensity spectra under the two ambient light conditions after calibration, from which we can clearly observe the intensity changing pattern between various video frames.

#### 4.3 Smartphone Screen Spectral

Through the pipeline of data preprocessing, we obtain clear fruit RGB video frames and NIR images for spectral reconstruction. However, since the light source of the system is the combination of the RGB LEDs rather than real monochromatic light, the previous spectral reconstruction model cannot be directly applied. The HSI images collected by the hyperspectral system can be formulated as  $I_{HSI} = E \cdot R \cdot S$ , where  $E$  is the emission intensity of the light source,  $S$  is the sensitivity of the CMOS detector of the HSI system, and  $R$  is the reflectance spectrum of the sample. Typically, the HSI system contains a slit to separate the light source into monochromatic light beams, thus we can regard  $E$  as an array of monochromatic light sources, i.e.,  $[E_{\lambda_1}, \dots, E_{\lambda_M}]$ . Comparing with the HSI system, the system separates the light source by modulating the screen display. Thus, the collected data can be formulated as:

$$I_{phone} = E_{screen} \cdot R \cdot S_p, \quad (4)$$

where  $S_p$  is the sensitivity of the smartphone cameras, which contains only three channels.  $E_{screen}$  is the screen-simulated monochromatic light sources, contains  $K$  monochromatic frames. The screen illumination has a fixed non-linear mapping to smartphone display sensors and monochromatic wavelength according to the CIE-RGB Color Space, i.e.,

$$E_{screen} = [\Phi(E_{RGB}, E_{\lambda_1}), \dots, \Phi(E_{RGB}, E_{\lambda_K})], \quad (5)$$

where  $3 \ll K < M$ . According to [73], the larger the number of spectral bands obtained, the more accurate the reconstructed curves are, theoretically. Thus, FruitPhone is expected to have better reconstruction performance than RGB-based solutions. However, since  $E_{screen}$  is different from the real monochromatic light sources  $[E_{\lambda_1}, \dots, E_{\lambda_K}]$ , we can not directly apply existing reconstruction algorithms on  $I_{phone}$ .

To solve this challenge, we propose a two-stage spectral reconstruction (SR) algorithm, as shown in Figure 12, which separates the screen spectral reconstruction problem into two tasks: spectral translation and spectral reconstruction. Firstly, we build a three-branch transformer structure to translate the screen spectra into monochromatic spectra with the same dimension, *i.e.*, to translate data from  $(x, y, 3, k)$  to  $(x, y, k)$ . This process mimics the inverse step of CIE-RGB color space translation process, thus we process the RGB channels in isolation. Secondly, we conduct the spectral reconstruction step on the translated spectra and combine it with the captured NIR images. This module focuses on enhancing the low spectral resolution caused by smartphone cameras.

**Spectral Translation.** The target of spectral translation is to translate the pseudo color spectra to the real absorption spectra at the same number of channels. We notice that a simulated monochromatic light can be taken as a mapping from a real monochromatic sensor to RGB LEDs, *i.e.*,  $E_{screen,\lambda} = \Xi(E_{RGB}, E_\lambda)$ . Thus, without changing the spectral dimension, we can realize the mapping of smartphone screen spectra to multi-spectral images. As shown in Figure 12(a), we achieve this by a three-branch transformer-based model. The shape of the input RGB data is  $(W, H, K, 3)$ , where  $K$  is the video frames that related to the screen monochromatic pattern and 3 is the R,G,B channels according to the smartphone camera sensitivity. Thus, we first separate the three channels and conduct translation separately, and then concatenate the translated features before input into the convolution embedding layer. This is because the spectral space translation process only relates to the screen light sources, manifested by the consecutive video frames. After that, we splice the output features together in order because each feature represents a spectrum over the specific wavelength range. The translation of each channel is related to all frames, which is very similar to the problem of two-language translation. Therefore, we refer to a transformer-based model, *i.e.*, SPECAT [69], as the backbone of our translator model (shown in Figure 12(b)). The translator contains several cumulative-attention blocks (CABs) within an efficient hierarchical framework to extract features from non-local spatial-spectral details. After translation, the concatenated features are input into a 1x1 convolution layer to interact information across the channels. The output size is  $(W, H, F)$ .

**Spectral Reconstruction.** We then utilize the state-of-the-art spectral reconstruction model, *i.e.*, MST++ [9], to increase the channel numbers of the translated spectra. It is worth noting that we introduce the NIR-band image during the spectral reconstruction stage to enhance the reconstruction accuracy in the NIR range, which contains rich absorption characteristics of the fruit sugar content. Similar to our spectral translation backbone, MST++ [9] utilizes a spectral-wise multi-head self-attention block to capture non-local self-similarity and long-range dependencies during the reconstruction process, which has shown strong capability on both RGB and MSI reconstruction tasks. We make some simple modifications on MST++ [9] to fit our reconstruction task. Firstly, we modify its input and output size to fit our dataset. Then, instead of using three single-stage spectral-wise transformer (SST) blocks, we use only one SST block since the increasing number of input channels simplifies the reconstruction difficulty.

**Shrinkage Loss.** Furthermore, we observe that the reconstruction model is not sensitive to hard samples, as few samples may have large reconstruction error, leading to extremely poor accuracy for sugar content regression. To solve this problem, we introduce a shrinkage loss function according to [39] to penalize the importance of easy training data. Specifically, we propose a modulating factor to re-weight the square loss to penalize easy samples only. The shrinkage loss function is defined as:

$$\mathcal{L} = \frac{L_{MSE}}{1 + \exp(a \cdot (c - L_{MRAE}))}, \quad (6)$$

where  $a, c$  are the hyper-parameters controlling the shrinkage speed and the localization respectively.  $L_{MSE}$  is the mean-square-error of the sample and  $L_{MRAE}$  is the mean relative absolute error that computes the pixel-wise disparity between all wavelengths of the reconstructed HSIs ( $\hat{Y}(\lambda)$ ) and ground-truth HSIs ( $Y(\lambda)$ ), which are

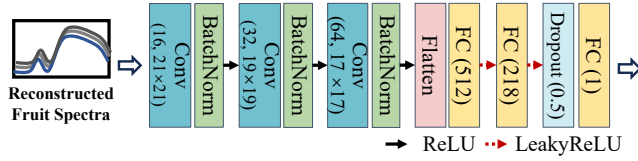


Fig. 13. The structure of the fruit sugar content regression model.

defined as:

$$L_{MSE} = \frac{1}{M} \sum_{i=1}^M [Y(i) - \hat{Y}(i)]^2, L_{MRAE} = \frac{1}{M} \sum_{i=1}^M \frac{|Y(i) - \hat{Y}(i)|}{Y(i)}. \quad (7)$$

By adding shrinkage penalty on easy samples, the reconstruction model can pay attention on minimizing the error of hard samples that have a large  $L_{MRAE}$ .

#### 4.4 Fruit Sugar Content Regression

Given the presence of redundant information in the hyperspectral imaging cubes, instead of using all pixels to train the regression model, we divide the HSI cube into several patches and calculate the mean spectrum for each patch for further regression analysis. In this work, each patch is a rectangle area of 8 pixels by 8 pixels. Additionally, we apply standard normal variate normalization to standardize the spectra to the same absorbance level, correcting for variations in optical path length and light scattering.

Different from previous fruit sugar content detection works that build regression models for each type of fruit [1, 29, 45, 63], we design a convolutional neural network (CNN) for all fruit samples. Since users may use FruitPhone to detect brix values for different fruits, building one generalizable model for all fruit types is more accessible than a fruit-dependent regression model. Figure 13 shows the architecture of our proposed CNN-based sugar content regression model. The model contains three convolutional layers to mimic the feature selection process of spectral analysis. We utilize a large kernel size for each convolutional layer, *i.e.*, 21, 19, and 17, to expand the broad receptive field so it can capture a wider range of context information in the neighboring wavelength bands. A three layer fully-connection structure is used to learn fruit sugar content related information from the extracted spectral features. We employ a dropout layer (rate=0.5) after the fully connected layer, a widely adopted configuration in dense layers, to mitigate overfitting risks [59]. It's worth noting that we utilize Leaky ReLU activation after the fully-connection layers to ensure no spectral features are discarded during backpropagation [8, 25]. The CNN structure simulates the spectrum analysis process of traditional stoichiometric methods, including feature extraction and nonlinear fitting.

Compared with traditional methods, the network based on deep learning has better fitting ability and generalization, making it more suitable for daily sugar content estimation scenarios. Table 1 shows the performance of traditional machine learning models<sup>1</sup> and our designed CNN model for fruit brix estimation. These traditional models have been utilized in previous fruit sugar content estimation works [1, 37, 63], showing good performance in single fruit sugar prediction. We compare these models using our collected hyperspectral images dataset, which contains 37 types of different fruits with brix values ranging from 3.1 ~ 23.1°Bx. We randomly select 20%

| Models          | $R^2$         | MAE           |
|-----------------|---------------|---------------|
| CARS+PLS        | 0.1947        | 3.2627        |
| SVR             | 0.2540        | 3.1933        |
| ELM             | 0.5211        | 2.6829        |
| 3-layer ANN     | 0.7457        | 1.8770        |
| <b>Ours CNN</b> | <b>0.8374</b> | <b>1.5205</b> |

Table 1. Performance of traditional machine learning models for fruit brix estimation.

<sup>1</sup>We select these models based on previous works. The partial least squares (PLS) model is widely used for fruit sugar content regression, such as in [1, 63]. The extreme learning machine (ELM) is a popular randomization-based learning algorithm that can be used for spectral analysis tasks [34]. The 3-layer ANN shares the same parameters with [37], demonstrating good accuracy in fruit sugar estimation.

of the hyperspectral images for testing and the use rest for training. From the results in Table 1, we can see that our proposed CNN model outperforms all existing solutions in the fruit sugar estimation task.

## 5 Implementation

We develop a proof of concept for FruitPhone on the Android platform, utilizing the commercially available Google Pixel4 XL due to its accessible near-infrared camera. To validate the extension capability of FruitPhone across different smartphones, we also implement the system on two other smartphones and evaluate the performance (in § 6.5.1).

**Smartphone Setups.** The application leverages the uniformly changed color patterns to scan samples while simultaneously recording with the front-facing camera and capturing NIR images. To ensure data stability across various ambient conditions, we maximize the screen intensity and maintain it at this level throughout the data collection process. Additionally, we set a capture time of 2 seconds, resulting in  $K = 45$  for each smartphone's data. This capture time enables users to complete the capture efficiently without excessive effort while obtaining as many frames as possible. A single run of the app on Google Pixel 4 XL consumes 2.945 mAh (0.079% of the battery capacity). The largest contributor is the prolonged screen usage (1.635 mAh, 55.5%), followed by the camera (1.2 mAh, 40.7%). The CPU usage was concentrated in the 1.2 GHz–2.8 GHz range, indicating that the application computing is relatively intensive. Considering the prolonged screen usage may contribute relative high power consumption, we also discuss the influence of shortening this capturing time and altering the uniform change pattern in § 6.5.2.

**Utilize Pipeline.** The reconstruction and sugar regression models are trained on a workstation. Once trained, these models are deployed to the mobile device for real-time inference. After the user capture a piece of new data, the captured frames undergo alignment and de-noising using the preprocessing methods outlined in Section 4.2. We utilize the SAM model to extract the region of interest (ROI) at the center of the fruit while discarding background information, thereby reducing the computational load. The SAM model processes both RGB and NIR images, which takes approximately 2.04 seconds. For data preprocessing, we enhance efficiency through vector calculations. The detrending algorithm is relatively time-consuming, requiring 0.2726 seconds per image, while black-and-white calibration takes only 0.014 seconds. Subsequently, the reconstruction model converts the collected data into hyperspectral data consisting of 138 channels within the 400 to 1000 nm range, which takes 0.1177 seconds for a single image sized  $50 \times 50$  pixels. The reconstructed hyperspectral data is then fed into the pre-trained sugar regression model for sugar content prediction, which requires an additional 0.1207 seconds per image. Finally, the results are communicated to the user, providing immediate feedback on the sugar content of the scanned samples. In total, each interference takes approximately 3 seconds.

## 6 PERFORMANCE EVALUATION

This section evaluates the performance of FruitPhone for brix estimation of various types of fruits. We validate the performance of FruitPhone by answering three key questions: 1) Does FruitPhone achieve better performance of sugar content detection than baselines? (§ 6.2) 2) Do the modules of the system effectively improve the performance? (§ 6.3) 3) Can FruitPhone achieve robust adulteration inspection under various experimental setups? (§ 6.4)

### 6.1 Study Setup

**6.1.1 Data Collection.** We prepare a total of 335 fruit samples from 37 different types or varieties purchased from a local supermarket, as shown in Table 2.<sup>2</sup> These fruits have diverse characteristics in terms of textures, colors,

<sup>2</sup>We open our processed dataset (335 hyperspectral images, RGB color spectra, near-infrared reference, and ground true brix values) on [https://drive.google.com/drive/folders/1jJQgOuB\\_yFYtod44AE5TEXKQXsivVbmj?usp=sharing](https://drive.google.com/drive/folders/1jJQgOuB_yFYtod44AE5TEXKQXsivVbmj?usp=sharing).

Table 2. Statistics of the fruit dataset, containing 37 fruit types and 335 samples.

| Category         | Fruit Type (Number of Varieties)                                                                                        | Sample | Brix Range ( $^{\circ}\text{Bx}$ ) |
|------------------|-------------------------------------------------------------------------------------------------------------------------|--------|------------------------------------|
| Berry            | Grapes (6), Tomato (2), Strawberry (1), Blueberry (1), Mulberry (1), Guvav (1), Kiwi (1), Passion Fruit (1), Pepino (1) | 180    | 3.1-23.1                           |
| Drupe and Kernel | Apples (4), Pear (2), Peach (1), Jujube (1), Loquat (1), Mango (1), Cheery (1), Persimmon (1)                           | 75     | 8.9-20                             |
| Citrus           | Mandarin (3), Orange (2), Lemon (2), Kumquat (1), Grapefruit (1)                                                        | 75     | 4.6-21.5                           |
| Cucurbits        | Longan (1)                                                                                                              | 5      | 11.4-16.3                          |

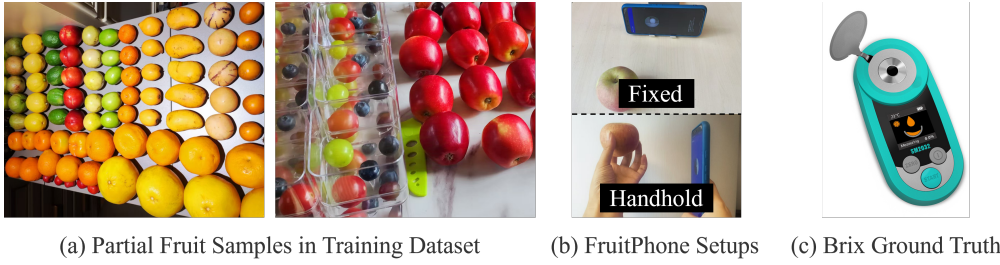


Fig. 14. Partial samples in the training dataset and ground truth devices used in evaluation.

shapes, sizes, and chemical compositions. The brix values range from  $3.1 \sim 23.1^{\circ}\text{Bx}$ , covering all levels of fruit quality. We also consider some thick-skin fruits like lemon and grapefruit, whose skin depth is about  $1.5\text{cm}$ . We prepare 5-20 samples for each fruit type. For some common fruits like apples or grapes, we also prepare several varieties whose physical and chemical characteristics are diverse. The fruits are purchased from several stores and at different times to ensure the diversity of freshness/ripeness levels. Figure 14(a) shows a portion of our fruit samples. We collect hyperspectral images using a Cubert FirefLEYE S185 Hyperspectral Imager [21] (about \$20,000) under two halogen lights at room temperature. Correspondingly, we capture the videos using the Google Pixel 4 XL smartphone with a capture time of 2 seconds and NIR images for each sample under the setup shown in Figure 14(b). The smartphone and fruit samples are placed on the desk during data collection process for quickly capturing. Only data in §6.4.5 are collected in the handheld setup. Additionally, all data are collected at room temperature, accompanied by an indoor light source from a fluorescent lamp and a small amount of sunlight. We align the data and extract the ROI for each fruit sample using processing methods depicted in §4.2, introducing a total of 335 paired HSI-phone data with a resolution of  $50 \times 50$  pixels. The ground truth sugar content of each fruit sample is detected by a portable but invasive digital refractometer, named SM20 Saccharimeter [44]. We juice each sample and then drop  $3\text{ml}$  into the sample bin of SM20 to measure the sugar content by analyzing the refractive index.

**6.1.2 Training Scheme.** To train the two-stage spectral reconstruction model, we split the images into  $32 \times 32$  patches with a 16-pixel stride. The batch size is 64, and the initial learning rate is set as  $10^{-4}$  with a cosine learning rate decay to  $10^{-6}$ . The hyperparameters  $a$  and  $c$  in the shrinkage loss are defined as 10 and 0.2 respectively, which means we will penalize samples whose  $L_{\text{MRAE}}$  is less than 0.2. The reconstruction model converges after 20,000 iterations. In addition, we split the image into  $8 \times 8$  pixel patches and calculate the average spectra of each patch to train the regression model. The model can quickly converge after 50 epochs with a batch size of 256. For both models, we utilize the Adam optimizer, which is the most common optimizer.

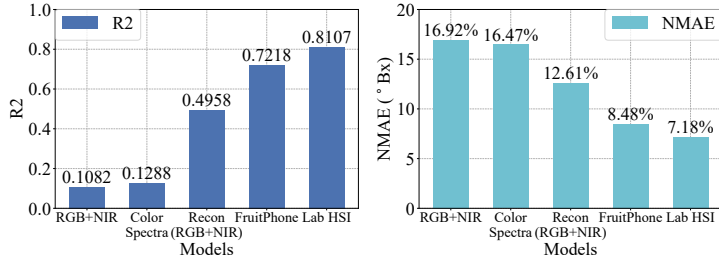


Fig. 15. Overall performance of FruitPhone and baselines on fruit sugar content regression.

|            | FruitPhone | Baseline |
|------------|------------|----------|
| MRAE       | 0.1803     | 0.2643   |
| RMSE       | 0.0981     | 0.1226   |
| PSNR       | 23.81      | 20.19    |
| Flops (G)  | 230.6      | 221.7    |
| Params (M) | 31.44      | 31.26    |

Table 3. Comparing the reconstruction performance and model capacity with the SOTA reconstruction baseline.

**6.1.3 Evaluation Metrics.** We use three commonly used metrics in hyperspectral imaging works to evaluate our spectral reconstruction model: mean relative absolute error (MRAE), root mean square error (RMSE), and peak signal to noise ratio (PSNR) [9, 33]. Among these, MRAE and RMSE evaluate the disparity between all wavelengths of the reconstructed and ground-truth HSIs. Since our target is to use the reconstructed spectra for fruit component analysis, we pay more attention to spectral-wise rather than spatial-wise errors. As for the brix regression task, we exploit R-squared ( $R^2$ ), root mean square error (RMSE), and normalized mean absolute error (NMAE) as evaluation metrics.

**6.1.4 Baselines.** We compare the performance of FruitPhone with the following schemes as baselines:

- **RGB + NIR Values** [7]. Since the first frame is illuminated by a white screen, which can provide spectral information across all wavelengths, we take the first frame of smartphone video data as the RGB input and utilize a multiple linear regression (MLR) model to predict the fruit sugar content. Additionally, considering the fruit sugar content has many absorption bands in the near-infrared range, we add NIR values as inputs to enhance this baseline.
- **Color Spectra** [58, 70]. Referring to previous works, we translate the captured video with  $k$  frames into a  $k$ -channel color spectrum, *i.e.*, calculate mean values across the RGB dimension. The NIR image is also added to this baseline to enhance its capability. We then utilize the extreme learning machine (ELM) to estimate the brix values, as ELM outperforms all other machine learning models in Table 1.
- **Recon (RGB + NIR)** [55]. We conduct spectral reconstruction on paired RGB and NIR images using MST++ [9]. The reconstructed data is transformed into one-dimensional spectra and used to predict brix values through our designed CNN-based regression model.
- **Lab HSI.** We input the HSI data collected by the expensive laboratory hyperspectral imager (*i.e.*, Cubert S185) into our CNN-based regression model to predict the brix values. This baseline is expected to have better performance than FruitPhone, due to the high quality of the input data.

## 6.2 Overall Performance

We evaluate the system's performance with 5-fold cross-validation on 335 samples. The data in the training set and test set are from completely different fruit samples.

**6.2.1 Brix Regression Performance.** We begin by evaluating the sugar content regression capability of FruitPhone and comparing it with baseline models. The overall prediction results are shown in Figure 15 - 17 and Table 4.

**Comparison with Baselines.** Figure 15 illustrates the NMAE and  $R^2$  of the prediction results respectively, revealing important findings. Firstly, FruitPhone achieves significantly better forecast results compared to all

Table 4. The sugar content estimation results of 39 types of food.

| Fruit            | RMSE   | NMAE   | Fruit            | RMSE   | NMAE   | Fruit           | RMSE   | NMAE   |
|------------------|--------|--------|------------------|--------|--------|-----------------|--------|--------|
| Apple (#1)       | 1.7156 | 5.70%  | Green Grape (#1) | 1.6847 | 7.80%  | Passion         | 3.1959 | 14.43% |
| Apple (#2)       | 2.2968 | 9.07%  | Green Grape (#2) | 1.3114 | 5.23%  | Peach (#1)      | 2.1872 | 10.39% |
| Apple (#3)       | 3.095  | 13.14% | Guava            | 0.9351 | 4.04%  | Peach (#2)      | 0.6102 | 2.34%  |
| Apple (#4)       | 3.339  | 14.03% | Jujube           | 1.2409 | 5.11%  | Pear            | 2.3218 | 8.74%  |
| Banana*          | 1.1740 | 4.97%  | Kiwi             | 1.8047 | 8.04%  | Pepino          | 1.958  | 7.7%   |
| Black Grape (#1) | 1.9671 | 8.79%  | Kumquat          | 7.0421 | 31.54% | Persimmon       | 3.1218 | 11.42% |
| Black Grape (#2) | 2.4378 | 9.01%  | Lemon (Green)    | 0.7925 | 2.90%  | Red Grape (#1)  | 2.1047 | 8.23%  |
| Blood Orange     | 1.3554 | 5.45%  | Lemon (Yellow)   | 1.9037 | 9.07%  | Red Grape (#2)  | 2.0841 | 10.42% |
| Blueberry        | 3.0260 | 10.45% | Longan           | 1.9262 | 8.12%  | Strawberry      | 1.0948 | 3.93%  |
| Cherry           | 2.6624 | 9.51%  | Loquat           | 4.308  | 15.78% | Sugar Orange    | 2.8539 | 12.3%  |
| Citrus           | 2.1050 | 9.55%  | Mango            | 2.6802 | 11.25% | Tomato (Red)    | 0.9653 | 4.12%  |
| Dekopon          | 3.1905 | 13.85% | Mulberry         | 2.6418 | 12.33% | Tomato (Yellow) | 3.6453 | 14.63% |
| Grapefruit       | 1.1010 | 4.66%  | Orange           | 1.3708 | 5.49%  | Watermelon*     | 2.4068 | 10.99% |

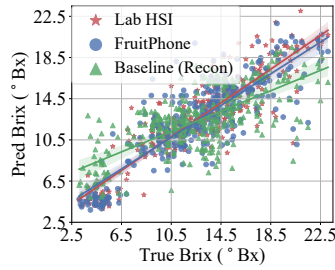


Fig. 16. FruitPhone's sugar content prediction results over all samples.

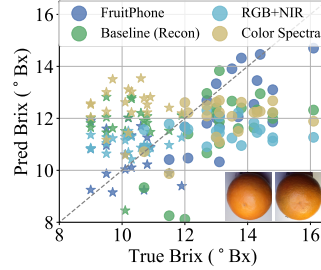


Fig. 17. Brix estimation performance on similar-look samples: orange (star) and blood orange (round).

|       | w/o De-noising | w/o Two-stage SR | w/o Shrink-age Loss |
|-------|----------------|------------------|---------------------|
| MRAE  | 0.2205         | 0.2864           | 0.2060              |
| RMSE  | 0.1257         | 0.1444           | 0.1224              |
| PSNR  | 21.96          | 18.66            | 22.23               |
| $R^2$ | 0.6430         | 0.1265           | 0.6301              |
| NMAE  | 9.56%          | 16.11%           | 9.60%               |

Table 5. Ablation study of FruitPhone in metrics of both reconstruction errors and brix regression performance.

baselines. Compared to the current SOTA method [55], the average  $R^2$  of FruitPhone increases by 0.2260 and the NMAE decreases by 4.12%. Even compared to a high-resolution and high-cost laboratory HSI system, FruitPhone is only inferior with an  $R^2$  value of 0.0891, demonstrating the effectiveness of the system. Secondly, we note that both the SOTA reconstruction solution and our FruitPhone show great improvement over raw data without reconstruction, which indicates the necessity of fine-grained spectral information when conducting in-depth chemical component analysis. Additionally, we observe that even with many more channels than RGB images, the color spectra achieve little improvement over RGB images. This is because although the color spectra mimic monochromatic light using RGB LEDs, they cannot be directly used as absorption spectra for component analysis. Consequently, it's quite valuable to conduct spectral reconstruction over the screen color spectra. Last but not least, from Figure 16, we can clearly observe that FruitPhone can estimate the brix values for fruits with various sugar content levels. Overall, FruitPhone performs comparably to laboratory hyperspectral imaging solutions and significantly outperforms state-of-the-art reconstruction-based methods.

**Errors Analysis across Fruit Characteristics.** Additionally, we carefully check the sugar content regression errors for each fruit and analyze the factors influencing these errors. Table 4 presents the regression results for

all fruit varieties. We analyze the potential factors from the aspects of shapes, sizes, colors, and peel thicknesses. **Firstly**, we find that the shape of the fruits—such as long strips (red grapes), triangles (strawberries, mangoes, pears), and spheres (apples, oranges)—do not significantly affect the prediction results. This is because we utilize our proposed detrend algorithm to mitigate the non-uniformity in spatial distribution caused by fruit shape. **Secondly**, we compare the influence of color on the system's predictions and the baseline results. Our dataset includes fruits that appear very similar but have significantly different brix levels, such as oranges and blood oranges. Despite their similarities in size, shape, color, and texture, their brix levels range from  $9^{\circ}\text{Bx}$  to  $16.1^{\circ}\text{Bx}$ . Figure 17 illustrates the prediction performance for 20 orange samples (star markers) and 20 blood orange samples (round markers). In this case, traditional colorimetric methods, like color spectra and RGB analysis, struggle to differentiate these visually similar samples. In contrast, our system reconstructs fine-grained hyperspectral images from RGB videos, allowing for in-depth spectral absorption analysis rather than relying solely on superficial colorimetry for assessing fruit content. **Thirdly**, our dataset includes fruit samples with medium or thick peels, such as grapefruits, lemons, and oranges, which have an approximate peel thickness of 1–1.5 cm. As the results shown in Table 4, the prediction performance for these fruits is comparable to that for thin-skinned fruits, like apples. This confirms the feasibility of the system in accurately predicting sugar content across different peel thicknesses. **Lastly**, to illustrate the generalizability of the system in large fruit types, we add verification results for two fruit types with significant shape differences from the original data set: bananas and watermelons. For each fruit type, we prepare 10 different samples, with sugar content ranging from  $4.7^{\circ}\text{Bx}$  to  $11.4^{\circ}\text{Bx}$ , and conduct tests using the system trained on the previous dataset with 37 fruit types. Table 4 presents the prediction results for these two fruit types, showing an RMSE of  $1.1740^{\circ}\text{Bx}$  for bananas and  $2.4068^{\circ}\text{Bx}$  for watermelons. The performance for these fruits aligns closely with that of the original dataset, further confirming the system's generalizability in predicting sugar content for larger fruit types.

**Errors Analysis across Fruit Types.** However, to our surprise, the prediction errors of fruits such as kumquats, yellow tomatoes, passions, and loquats are even larger. These fruits do not have obvious commonalities in physical structure and the spectral reconstruction errors are not significant. By comparing with the results of laboratory-level HSI predictions, we find that even with more accurate HSI predictions, the results of these fruit varieties are not ideal, which is related to the chemical interfering substances they contained. Taking kumquats as an example, we observe that the system's prediction for kumquats is very poor, with RMSE exceeding  $7^{\circ}\text{Bx}$  (compared to  $4.8707^{\circ}\text{Bx}$  RMSE by laboratory-level HSI). This is because the high concentration of limonene on its surface generates a strong absorption peak at 1380 nm, which may cover the absorption characteristics of glucose [36, 74]. To solve this problem, fruit-dependent model may be needed for specific fruit types. After two-stage reconstruction, we can identify the fruit type by object detection method and then apply corresponding fruit model.

**6.2.2 Spectral Reconstruction Error.** Table 3 shows the reconstruction error of FruitPhone compared to the state-of-the-art spectral reconstruction model, MST++ [9]. The baseline model uses paired RGB and NIR image as input, which only has four channels. While FruitPhone employs screen spectra with 45 frames along with the paired NIR image, totaling 46 input channels. The table clearly shows a significant improvement in all three metrics for FruitPhone over the baseline. This enhancement is attributed to the higher number of input channels in FruitPhone, allowing for more precise and accurate reconstruction relationships, consistent with our feasibility study findings. Additionally, we note that both the baseline model and our reconstruction model achieve relatively low PSNR values. This is due to the fact that we only input the central area of the fruit images into the reconstruction models, which lacks sufficient spatial characteristics for the models to learn from. However, since we rely solely on spectral features for brix estimation, the poor spatial reconstruction quality does not affect the overall performance of our system.

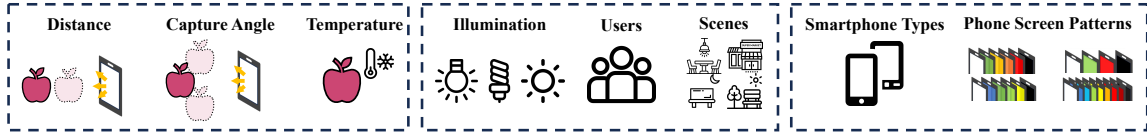


Fig. 18. Illustration of differing conditions we have considered, including differences in fruit sample states, environmental factors, and smartphone configurations.

**6.2.3 Model Overhead.** In addition, we present the floating-point operations per second (FLOPS) and total parameters for both our two-stage reconstruction model and the baseline in Table 3. A comparison with the baseline reveals that FruitPhone delivers significantly improved reconstruction performance without a substantial increase in computational load. Although FruitPhone employs a complex network architecture, we have reduced the number of transformer blocks from the original MST++ model. As a result, our proposed two-stage spectral reconstruction achieves enhanced performance on the screen spectral reconstruction task without imposing excessive computational demands.

### 6.3 Ablation Study

We then investigate the designs of FruitPhone by conducting ablation studies to demonstrate the effectiveness of each design. Table 5 presents the results of our ablation study.

**(1) W/O De-noising Module.** Compared to the system with a de-noising module, *i.e.*, spatial detrending and white-black calibration, the model without de-noising module has a decrease in both reconstruction and brix regression accuracy. This is because ambient interference and distance shift are unique for each sample, and if these noises are not eliminated, the generalization error between the training samples and the test samples will be caused.

**(2) W/O Shrinkage Loss.** We can see that without shrinkage loss, the spectral reconstruction error only slightly increases but the brix regression performance decreases obviously. This is because the shrinkage loss can reduce the reconstruction error of a small number of difficult samples in the reconstructed model, which does not significantly lower the overall reconstruction error. However, these "hard samples," which exhibit large reconstruction errors, greatly affect the sugar estimation task. For instance, the model without shrinkage loss performs even worse in brix estimation than the model lacking a de-noising module, despite having a smaller reconstruction error. This underscores the importance of incorporating shrinkage loss.

**(3) W/O Two-Stage Spectral Reconstruction.** Additionally, we remove the spectral translation step and directly input paired video frames and NIR image into the state-of-the-art SR model, *i.e.*, MST++. Unlike the baseline, which has only four input channels, the model in this ablation study treats the 45 frames of the RGB video as 45 spectral channels. From the results shown in Table 5, we can find that the system introduces a terrible reconstruction performance, even worse than the baseline with only four channels. This is because the frames and spectral channels of the RGB video are not physically equivalent. When video frames and near-infrared images are used directly for reconstruction, the model struggles to learn the mapping between the input spectra and the hyper-spectra. This further emphasizes the necessity of converting the screen spectrum into the real spectrum space prior to spectral reconstruction.

### 6.4 Differing Conditions

We then evaluate FruitPhone under various experimental and environmental conditions and compare it with the baseline using SOTA reconstruction solutions. Figure 18 shows all conditions that we considered, including the variation of the distance from the fruits, angles, temperatures, ambient illuminations and users. We train

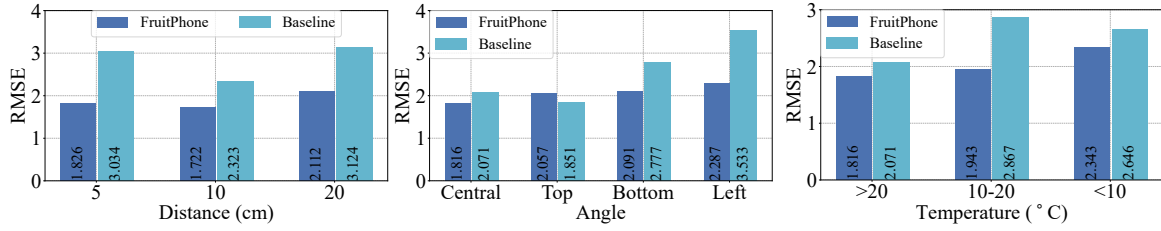


Fig. 19. Brix regression error of FruitPhone and the SOTA reconstruction solution under various capturing distance, angle, and sample temperatures.

both the reconstruction and regression models using the dataset collected under default settings in §6.1.1. To assess the system’s robustness, we use 16 fruit samples from four typical fruit types, *i.e.*, apple, orange, tomato, and black grape. We collect data under various conditions to evaluate FruitPhone’s performance across these different scenarios.

**6.4.1 Varying Data Capturing Distance.** Although we recognize the system to be used at a capturing distance of 10 cm, it’s difficult for users to exactly meet this. Thus, we investigate the performance of FruitPhone by capturing the sample at different distances to the smartphone, including 5 cm, 10 cm, and 20 cm. The results are depicted in Figure 19. We can find that the performance of FruitPhone is consistently better than the baseline. However, with the sample getting too far from the smartphone camera, the performance of FruitPhone slightly decreases. This is because when the shooting distance is too far, the intensity of the light hitting the surface of the sample is weak, so the quality of the collected data is poor, resulting in a decline in performance. One idea to solve this problem is to increase the luminous intensity of the mobile phone screen, but this may lead to an increase in the energy consumption of the mobile phone. Therefore, we recommend that the use distance of FruitPhone be within 15cm, which is very easy to achieve.

**6.4.2 Varying Data Capturing Angle.** Users may capture the samples at different angles causing various light field distributions on the fruit surface. Thus, to evaluate the robustness of FruitPhone under various capturing angles. We consider the central area and location relative to the circumference of the smartphone, each containing approximately 20° angles with the center of the smartphone’s front-camera. Figure 19 demonstrates the results, which indicate that FruitPhone is robust to various capturing angles. This is because we utilize a spatial de-trending approach to eliminate the unevenness of the light field distribution. Therefore, the processed data can produce stable spectral data in the face of incident light from different angles.

**6.4.3 Varying Fruit Temperature.** Temperature is a common factor that may influence the spectral features of fruit samples. In real cases, some fruits are refrigerated at less than 15°C to keep them freshness. Thus, it’s necessary to investigate the robustness of FruitPhone when predicting sugar content values from fruit samples under different temperatures. We simulated the possible temperature in the real world and placed the samples in the refrigerator for 5 hours (about 10-20 degrees) and 18 hours (below 10 degrees). Figure 19 shows the results of FruitPhone measuring the sugar content of fruit samples at different temperatures, from which it can be seen that FruitPhone is more adaptable to different temperature conditions than the baseline. Even if the sample is refrigerated for a long time, the performance is relatively stable.

**6.4.4 Varying Ambient Illumination.** Ambient light can interfere with the readings of smartphone cameras, especially the RGB camera. Considering the emission spectrum of various light sources is different, which may

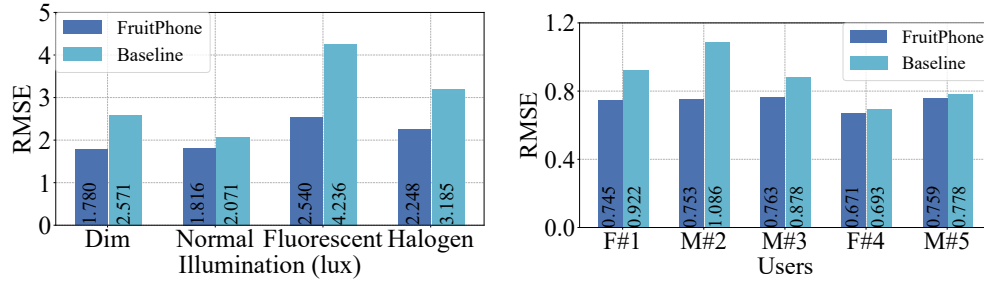
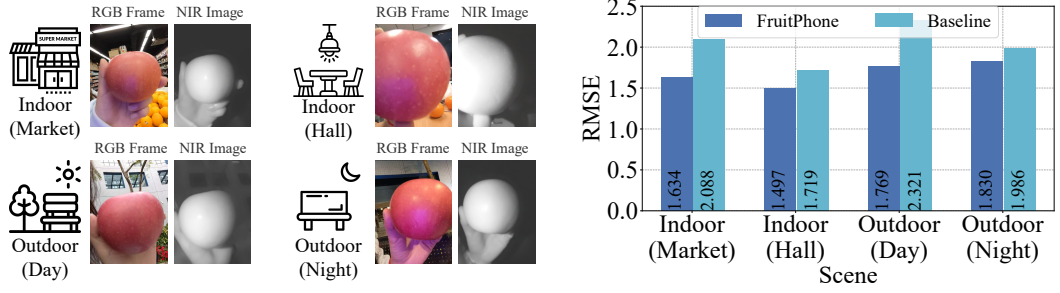


Fig. 20. Brix regression performance of FruitPhone and the SOTA reconstruction solution for different ambient illumination conditions and users operating FruitPhone.

bring diverse influence to the smartphone camera, we investigate FruitPhone’s robustness under various realistic lighting conditions, including different intensities of light in indoor environments, as well as two lamp types. From the results presented in Figure 20, we can observe that, FruitPhone’s performance under different light conditions is stable. This benefits from the white-black calibration step that subtracts the ambient interference during the data preprocessing pipeline. Additionally, we find that FruitPhone’s performance consistently overwhelms the baseline in all light conditions.

**6.4.5 Varying Users.** Besides the shooting distance and angle that can be quantitatively evaluated, there are many user habits, such as jitter, rotation, etc. To assess the system’s usability across different users, we invite five participants (two females and three males) to use FruitPhone implemented on the Google Pixel 4 XL smartphone. Before data collection, participants are briefed on FruitPhone’s operation and provided with recommended shooting parameters (as detailed in § 6.1.1). Each user captures images of two fruit types (apples and oranges) in an indoor environment at room temperature (17–23°C), with five samples per fruit. To ensure data reliability, every sample is scanned twice. Figure 20 shows the results for all users. We can see that FruitPhone does not vary much in performance from user to user. We find that user#4 has less hand movement during data collection, so the results are slightly better than those of other users. We can increase the alignment of video frames during data processing to eliminate the effect of jitter during video shooting.

**6.4.6 Varying Scenes.** Although considering several factors separately, real scenarios can be uncontrollable with compounding factors, including both sample angle, distance, and user jitter, as well as ambient conditions. Therefore, we test FruitPhone in four uncontrolled real-world scenes: (1) the supermarket with bright indoor fluorescent illumination, (2) the dining hall with both indoor light and sunlight from windows, (3) the outdoor park during the day with bright sunshine, (4) the outdoor park during the night with moonlight and street lamp. These scenarios are selected to represent different combinations of environmental illuminance (500–7500 lux), temperature and humidity. We analyze three types of fruits (apples, oranges, and citrus), which have nine different brix values (8.6–14.4 °Bx), and each type of fruit is measured twice. The samples are refrigerated for different durations and then exposed to the four scenarios, leading to diverse sample temperatures. As shown in Figure 21, FruitPhone maintains a relatively stable performance of  $RMSE < 2.0^{\circ}Bx$  in all scenarios. It is worth noting that through our black and white calibration algorithm, the increase of error in high-illumination scenarios ( $>7500$  lux) is mitigated, limiting RMSE to  $1.92^{\circ}Bx$  under extreme light interference. In the outdoor night scene, where samples are exposed to long-term refrigeration (10 hours at 4°C), the results show a slightly higher deviation ( $RMSE=1.98^{\circ}Bx$ ). This might be due to the change in near-infrared absorption characteristics at suboptimal temperatures. We plan to address this recognized limitation through temperature-compensated spectral modeling.



(a) The same fruit sample captured under various scenes with diverse compounding factors.

(b) The RMSE of FruitPhone and the baseline for brix estimation under various scenes.

Fig. 21. Brix regression performance of FruitPhone and the SOTA reconstruction solution under real-world scenes with uncontrolled compounding factors, *e.g.*, angle, distance, temperature, user jitter, and ambient illumination.

in future work. Importantly, under typical supermarket conditions (20-25°C, 500-2500 lux), FruitPhone achieves stable accuracy (RMSE=1.12°Bx), confirming its practicality in real-world fruit selection scenarios.

## 6.5 Further Discuss

Since the spectral reconstruction is related to the smartphone hardware parameters, such as screen light sources and camera sensitivity. We further discuss the generalization of the system to different smartphone setups. As shown in Figure 18, we validate FruitPhone on different smartphones and various screen illumination patterns.

**6.5.1 Varying Smartphones.** Different smartphones have diverse hardware parameters, such as screen types, RGB camera sensitivity, and video processing methods. These variations may impact the data quality of the collected video frames. This issue can be addressed through model fine-tuning or transfer learning. Given that screen types may be the most significant factor influencing data quality, we choose three smartphones with varying screen technologies, *i.e.*, LCD, OLED, and AMOLED and test the generalization capability cross different smartphones. Table 6 lists the parameters of three smartphones, which are different in both screen and camera parameters. We use the three smartphones to collect video data for 135 fruit samples. Both the reconstruction and regression models are pre-trained using 80% of the data collected by Google Pixel 4 XL, and then fine-tuned using the corresponding data collected by the other phone. We freeze all the model except the last layer of each channel translator and the last layer of the reconstruction module. Note that since the OnePlus and IQOO smartphones can not access of the NIR camera, for fair comparison, we only use RGB video for training and fine-tuning of the reconstruction model. Table 6 shows the results after transfer learning. We can observe that after fine-tuning, the reconstruction errors on the new smartphone only slightly increase, which does not influence the sugar regression performance too much. Additionally, we notice that the OnePlus smartphone exhibit the largest reconstruction error. This may because it applies a different video processing method to the other two smartphones, which will adjust the color and brightness of the video frame by frame, generating fluctuation in the collected video frames. This issue could be mitigated by incorporating an additional processing step into the data collection procedure for the OnePlus smartphone.

**6.5.2 Varying Screen Monochromatic Light Patterns.** We then discuss how the number of input color spectra channels affects the spectral reconstruction performance. We evaluate the two-stage spectral reconstruction using phone spectra with different channel numbers, *i.e.*, 3, 15, 30, and 45. The model is trained on 80% of

Table 6. Comparing the parameters across three smartphones and FruitPhone’s performance on them after fine-tune.

|                  | Screen Type  | Screen Resolution | Refresh Rate | Camera Resolution | HDR Support | MRAE   | RMSE   | $R^2$ | NMAE  |
|------------------|--------------|-------------------|--------------|-------------------|-------------|--------|--------|-------|-------|
| Google Pixel 4XL | OLED         | 3040×1440         | 90Hz         | 8MP               | HDR10       | 0.1954 | 0.1032 | 23.15 | 8.90% |
| OnePlus 8 Pro    | Fluid AMOLED | 3168×1440         | 120HZ        | 16MP              | HDR10+      | 0.2120 | 0.1476 | 20.26 | 9.97% |
| IQOO Z9X         | LCD          | 2400×1080         | 120HZ        | 16MP              | HDR10       | 0.2068 | 0.1353 | 20.79 | 9.25% |

|      | Numbers |        |        |        |
|------|---------|--------|--------|--------|
|      | 3       | 15     | 30     | 45     |
| MRAE | 0.1518  | 0.1461 | 0.1451 | 0.1441 |
| RMSE | 0.0844  | 0.0838 | 0.0836 | 0.0781 |

Table 7. Spectral reconstruction errors of FruitPhone when using various numbers of input channels.

|                | Recon  |        | Regression |        |
|----------------|--------|--------|------------|--------|
|                | MRAE   | RMSE   | $R^2$      | RMSE   |
| Uniform        | 0.1557 | 0.0767 | 0.8713     | 1.7219 |
| Characteristic | 0.1563 | 0.0767 | 0.8873     | 1.5908 |

Table 8. Spectra reconstruction and brix regression performance of FruitPhone using various screen patterns.

the collected 200 paired spectra and tested on the other 20% of the data. Table 7 shows the MRAE and RMSE of the reconstructed spectra on the test dataset. We can observe a decreasing trend of RMSE loss with more channels used for training the model. Furthermore, we also evaluate how monochromatic pattern influence the usability of FruitPhone, by modifying the screen color pattern. We select a series of characteristic wavelengths according to the wavelength selection methods. Then the light mode of the phone screen is set according to the selected characteristic wavelength. Therefore, the phone screen will illuminate more time in the sugar-related wavelengths. Table 8 shows the reconstruction performance of FruitPhone using uniform pattern and characteristic pattern from the same 5-fold cross validation on 200 fruit samples. The results are valuable for further discussion, as we find that the characteristic pattern has slightly larger reconstruction errors, but achieves better brix regression performance. This is because the HSI data has redundancy between bands, and not all bands are useful when applied to the sugar regression task. Therefore, although the overall reconstruction error is larger, the reconstructed spectrum performs better on the bands related to the sugar degree, so there is a chance to obtain better prediction results. However, in order to preserve the potential of FruitPhone to extend to more detection indicators in this paper, we use an unfiltered uniform spectrum as input.

## 7 RELATED WORK

In this section, we briefly review existing works related to portable fruit quality detection, smartphone-based spectral systems and spectral reconstruction technology.

### 7.1 Fruit Quality Detection

In recent years, various research solutions have emerged to facilitate the detection of fruit quality, particularly focusing on parameters like sweetness, often measured in brix. This is crucial for consumers who want to ensure they are purchasing high-quality produce.

Many studies have explored the use of image processing and machine learning algorithms to analyze fruit characteristics, such as color, texture, and shape, to determine ripeness and freshness [31, 48, 52]. For instance,

some apps utilize deep learning techniques to classify fruit quality based on user-uploaded photos [52]. However, image-based solutions can only access superficial characteristics which is not sufficient for fruit in-depth quality assessment, e.g., sugar content, ripeness, or acidity detection [38, 72]. In addition, some research tries to explore other sensing modalities like ultrasound [6, 17], radio frequency [1, 37] to detect fruit quality parameters in a non-invasive manner. Although they can obtain accurate accuracy, these systems require particular signal transmission and receiving systems that cannot be easily supported by accessible devices.

Spectrometers have gained attention as a more direct method for assessing fruit quality. Since the spectrometer can penetrate the fruit deeper and obtain fine-grained spectral characteristic of the components inside the fruit, it allows users to measure parameters like sugar content and acidity more accurately. For instance, some portable spectrometers, such as SCiO [54] and the NIR handheld spectrometer [16], can provide real-time analysis of fruit quality parameters. While they offer high accuracy, their cost can be prohibitive for average consumers, often exceeding several hundred dollars. Some recent studies try to achieve smartphone-based spectral imaging systems to obtain high accuracy with low-cost, like MobiSpectral [55]. The key idea is to reconstruct full spectral information from smartphone captured three channel RGB/NIR images. However, due to the low quality of smartphone data, the spectra reconstructed by these methods perform poorly in accuracy and stability, and can only support roughly qualitative classification tasks, such as the classification of organic/inorganic fruits.

## 7.2 Smartphone-based Spectral Systems

Smartphones have become effective tools for spectral analysis due to their integrated cameras and optical sensors, such as complementary metal-oxide semiconductor (CMOS) sensors. Smartphone-based spectral systems can be categorized into three main types: colorimeters, spectrometers, and fluorimeters [14].

First, the individual color outputs of smartphones can function as colorimeters, enabling applications like water quality assessment [60], air quality monitoring [68], and bioanalytical testing [62]. Additionally, incorporating optical dispersion components, such as diffraction gratings, can enhance the spectral resolution of smartphone data [42, 43, 53]. Some studies have also developed fluorimeters using the smartphone's CMOS camera [11, 13, 46]. However, these approaches often require additional hardware, which can limit accessibility and reduce user willingness to engage with the system.

Recent research has focused on achieving optical analysis using unmodified smartphones [23, 55, 58], leveraging data processing algorithms to improve data quality. For instance, He et al. [23, 24] employed a spectral reconstruction algorithm with a standard smartphone to perform hyperspectral imaging for analyzing skin morphology and monitoring hemodynamics. MobiSpectral [55] achieved hyperspectral imaging for classifying organic and inorganic fruits using only the built-in RGB and near-infrared cameras of smartphones. While these methods do not require additional hardware, they reconstruct hyperspectral images directly from RGB data, which limits accuracy and complicates fine-grained detection tasks. Additionally, Song et al. [57, 58] proposed using smartphone screens to generate spectra for food authentication, sharing a similar concept with this work. However, the spectra collected under screen illumination differ physically from real spectra, making them unsuitable for direct qualitative object detection.

## 7.3 Spectral Reconstruction Algorithms

Hyperspectral imaging (HSI) is a powerful technology that combines spectroscopy with imaging capabilities. However, it faces challenges related to portability and affordability for most users. Recent efforts have focused on developing computational methods to reconstruct full hyperspectral data from limited spectral measurements, thereby reducing the reliance on expensive and bulky hardware. Some approaches involve designing specific reconstruction algorithms that rely on hand-crafted hyperspectral priors [3, 20, 35]. These methods often require specialized expertise and involve numerous parameters, making them difficult to apply in broader contexts.

Additionally, deep learning techniques, particularly convolutional neural networks (CNNs), have been utilized to learn the mapping from RGB images to hyperspectral images [2, 33, 56, 66, 67, 75]. For instance, Li et al. [33] integrated spectral response priors into an adaptive weighted attention network. More recently, transformer-based architectures have demonstrated superior performance over CNNs in spectral reconstruction tasks. MST++ [9] introduced a spectral-wise multi-head self-attention block to capture non-local self-similarity and long-range dependencies during the reconstruction process. Yao et al. [69] proposed a spatial-spectral cumulative-attention transformer designed for high-resolution hyperspectral image reconstruction.

However, these methods primarily focus on reconstructing high-resolution hyperspectral images from compressive spectral images, such as RGB and multispectral images, which exhibit a strong correlation with hyperspectral data. In the context of FruitPhone, the limitations of smartphone cameras and screen light sources result in spectral data that is not only compressed but also subject to color transformation. Consequently, previous reconstruction methods cannot be directly applied.

## 8 DISCUSSION

In this section, we shall discuss the limitations and future extensions of FruitPhone.

**Different Smartphone Types.** In theory, the smartphone-based spectral reconstruction is intrinsically linked to the parameters of smartphone hardware. This paper demonstrates good performance of FruitPhone on a Google Pixel 4 XL smartphone. However, variations in screen types among different phones (such as LCD and OLED), camera response, and image processing algorithms can result in disparate data readings, thereby impacting the reconstruction performance. Among these factors, differences in screen characteristics (including the central wavelength of RGB LEDs and sensor array displacement) are most pronounced and likely to affect data quality. In § 6.5.1, we have shown that through fine-tuning with a small data set, FruitPhone can achieve comparable reconstruction errors and sugar regression accuracy across different smartphones. In addition to the fine-tuning solution, we envision the potential for other unsupervised generalization methods that could facilitate portability between new smartphones. For instance, we could develop a normalization algorithm that translates images from various devices to conform to a standard format prior to their input into a unified reconstruction model. We consider this an avenue for future exploration.

**Non-NIR Smartphones.** Although FruitPhone has achieved a validation accuracy comparable to that of professional devices, we acknowledge that FruitPhone cannot be universally applied to all mobile phones. This limitation arises because FruitPhone utilizes the NIR sensor present in the mobile phone during the analysis process, but not all smartphones are equipped with this sensor. The evaluation in Table 6 shows that sugar prediction accuracy decreases on non-NIR smartphones, as the near-infrared band (700-1100 nm) provides crucial spectral features of chemical bonds like C-H and O-H. To improve near-infrared detection in these devices, we can optimize both hardware and algorithms. For hardware, integrating affordable NIR sensors into mobile phones is a possibility [12, 28]. For algorithms, we can use the correlation between spectral bands and information from the visible light band to infer near-infrared spectral features for ingredient prediction [5, 49].

**Thick-Skinned Fruit.** Analyzing thick-skinned fruits presents challenges due to the limited penetration depth of visible and near-infrared light. In this work, our dataset contains fruit samples with varying peel thickness ranging from 0.01 to 1.5 cm. Previous studies have examined the penetration depth of light in the 500-1000 nm can reach most fruits, including thin-skinned fruits like apples [32, 51] and medium-skinned fruits such as citrus [15], as well as watermelons [4], can be analyzed non-invasively using near-infrared light. However, for thick-skinned fruits like pomelo and pomegranate, this light is insufficient to penetrate the skin, resulting in inaccurate predictions regarding the internal quality and nutrient content of the fruit [15]. One potential solution is to investigate the relationship between the nutrient content of the peel and the pulp by studying biochemical

pathways and nutrient transfer mechanisms. This analysis could help predict the nutrient content of the flesh based on the spectral absorption characteristics of the skin.

**Extension Applications.** Given that FruitPhone enables more accurate spectral reconstruction, it facilitates extensive analysis tasks that require fine-grained spectral information, such as adulteration detection and nutritional analysis. By incorporating additional training data to enhance the spectral reconstruction model and deploying diverse application-specific regression models, FruitPhone has the potential to transform smartphones into new spectral analysis workstations.

## 9 CONCLUSION

This paper introduces FruitPhone, a smartphone-based spectral imaging system that can accurately quantify the sugar content of various fruits. By leveraging the smartphone screen to simulate multiple monochromatic light sources and employing a robust two-stage spectral reconstruction algorithm based on transformer modules, FruitPhone successfully enhances spectral resolution and reliability. Additionally, FruitPhone addresses environmental variability and spatial inconsistencies by using a white-black calibration and detrending algorithm for high-dimensional spectral images. Our results demonstrate that FruitPhone achieves a high degree of accuracy, closely rivaling expensive desktop spectrometers while being more affordable and user-friendly. Ultimately, FruitPhone empowers users to effortlessly and accurately evaluate fruit quality and sugar content in real-time, thereby enhancing their purchasing decisions and promoting healthier eating habits.

## Acknowledgments

We are grateful to all the anonymous reviewers for their insightful comments. This research is supported in part by RGC under Contract CERG 16206122, 16204523, 16205824, AoE/E-601/22-R, SRFS2425-6S05 and Contract R8015.

## References

- [1] Sayed Saad Afzal, Atsutse Kludze, Subhajit Karmakar, Ranveer Chandra, and Yasaman Ghasempour. 2023. AgriTera: Accurate Non-Invasive Fruit Ripeness Sensing via Sub-Terahertz Wireless Signals. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.
- [2] Aitor Alvarez-Gila, Joost Van De Weijer, and Estibaliz Garrote. 2017. Adversarial networks for spatial context-aware spectral image reconstruction from RGB. In *Proceedings of the IEEE international conference on computer vision workshops*. 480–490.
- [3] Boaz Arad and Ohad Ben-Shahar. 2016. Sparse recovery of hyperspectral signal from natural RGB images. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part VII 14*. Springer, 19–34.
- [4] Edwin R Arboleda, Kimberly M Parazo, and Christle M Pareja. 2020. Watermelon ripeness detector using near infrared spectroscopy. *Jurnal Teknologi dan Sistem Komputer* 8, 4 (2020), 317–322.
- [5] Masoomah Aslahishahri, Kevin G Stanley, Hema Duddu, Steve Shirliffe, Sally Vail, Kirstin Bett, Curtis Pozniak, and Ian Stavness. 2021. From RGB to NIR: Predicting of near infrared reflectance from visible spectrum aerial images of crops. In *Proceedings of the IEEE/CVF international conference on computer vision*. 1312–1322.
- [6] TS Awad, HA Moharram, OE Shaltout, DYMM Asker, and MM Youssef. 2012. Applications of ultrasound in analysis, processing and quality control of food: A review. *Food research international* 48, 2 (2012), 410–427.
- [7] Jayanta Kumar Basak, Bolappa Gamage Kaushalya Madhavi, Bhola Paudel, Na Eun Kim, and Hyeon Tae Kim. 2022. Prediction of total soluble solids and pH of strawberry fruits using RGB, HSV and HSL colour spaces and machine learning models. *Foods* 11, 14 (2022), 2086.
- [8] Archisman Bhattacharjee and Pawan Bharadwaj. 2025. Coherent Spectral Feature Extraction Using Symmetric Autoencoders. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (2025).
- [9] Yuanhao Cai, Jing Lin, Zudi Lin, Haoqian Wang, Yulun Zhang, Hanspeter Pfister, Radu Timofte, and Luc Van Gool. 2022. Mst++: Multi-stage spectral-wise transformer for efficient spectral reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 745–755.
- [10] Gregory A Carter and Richard L Miller. 1994. Early detection of plant stress by digital imaging within narrow stress-sensitive wavebands. *Remote sensing of environment* 50, 3 (1994), 295–302.

- [11] Arnold Chen, Royal Wang, Candace RS Bever, Siyuan Xing, Bruce D Hammock, and Tingrui Pan. 2014. Smartphone-interfaced lab-on-a-chip devices for field-deployable enzyme-linked immunosorbent assay. *Biomicrofluidics* 8, 6 (2014).
- [12] Soo Chung, Lane E Breshears, and Jeong-Yeol Yoon. 2018. Smartphone near infrared monitoring of plant stress. *Computers and Electronics in Agriculture* 154 (2018), 93–98.
- [13] Ahmet F Coskun, Justin Wong, Delaram Khodadadi, Richie Nagi, Andrew Tey, and Aydogan Ozcan. 2013. A personalized food allergen testing platform on a cellphone. *Lab on a Chip* 13, 4 (2013), 636–640.
- [14] Sarah Di Nonno and Roland Ulber. 2021. Smartphone-based optical analysis systems. *Analyst* 146, 9 (2021), 2749–2768.
- [15] Yiqing Dong, Guorong Du, Liwen Jiang, Yang Shan, and Pao Li. 2023. A new method for evaluating the penetration ability of near infrared diffuse reflectance light to fruit peel with chemometrics. *Vibrational Spectroscopy* 129 (2023), 103599.
- [16] Felix. 2024. F-750 Produce Quality Meter. <https://felixinstruments.com/food-science-instruments/nir-spectroscopy/f-750-produce-quality-meter/>
- [17] Mahmoud Soltani Firouz, Ali Farahmandi, and Soleiman Hosseinpour. 2019. Recent advances in ultrasound application as a novel technique in analysis, processing and quality control of fruits, juices and dairy products industries: A review. *Ultrasonics sonochemistry* 57 (2019), 73–88.
- [18] Food and Agriculture Organization. 2021. The State of Food and Agriculture. <http://www.fao.org/publications/sofa/2021/en/> Accessed: 2023-10-01.
- [19] Qiang Fu, Matheus Souza, Eunsue Choi, Suhyun Shin, Seung-Hwan Baek, and Wolfgang Heidrich. 2024. Limitations of Data-Driven Spectral Reconstruction—An Optics-Aware Analysis. *arXiv preprint arXiv:2401.03835* (2024).
- [20] Ying Fu, Yongrong Zheng, Lin Zhang, and Hua Huang. 2018. Spectral reflectance recovery from a single RGB image. *IEEE Transactions on Computational Imaging* 4, 3 (2018), 382–394.
- [21] Cubert GmbH. 2023. FIREFLYEYE 185. <https://www.cubert-hyperspectral.com/products/fireflye-185>.
- [22] Mayank Goel, Eric Whitmire, Alex Mariakakis, T Scott Saponas, Neel Joshi, Dan Morris, Brian Guenter, Marcel Gavrilu, Gaetano Borriello, and Shwetak N Patel. 2015. HyperCam: hyperspectral imaging for ubiquitous computing applications. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 145–156.
- [23] Qinghua He, Wanyu Li, Yaping Shi, Yi Yu, Wenqian Geng, Zhiyuan Sun, and Ruikang K Wang. 2023. SpeCamX: mobile app that turns unmodified smartphones into multispectral imagers. *Biomedical Optics Express* 14, 9 (2023), 4929–4946.
- [24] Qinghua He and Ruikang Wang. 2020. Hyperspectral imaging enabled by an unmodified smartphone for analyzing skin morphological features and monitoring hemodynamics. *Biomedical optics express* 11, 2 (2020), 895–910.
- [25] Haiyan Hu, Qianyi Huang, and Qian Zhang. 2023. Babynutri: a cost-effective baby food macronutrients analyzer based on spectral reconstruction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 1 (2023), 1–30.
- [26] Haiyan Hu, Qian Zhang, and Yanjiao Chen. 2022. NIRSCam: A mobile near-infrared sensing system for food calorie estimation. *IEEE Internet of Things Journal* 9, 19 (2022), 18934–18945.
- [27] Food Marketing Institute. 2021. Consumer Trends in Food Quality. <https://www.fmi.org/research-and-reports> Accessed: 2023-10-01.
- [28] Kacie Kaile and Anuradha Godavarty. 2019. Development and validation of a smartphone-based near-infrared optical imaging device to measure physiological changes in-vivo. *Micromachines* 10, 3 (2019), 180.
- [29] Sang-Yeon Kim, Suk-Ju Hong, Eungchan Kim, Chang-Hyup Lee, and Ghiseok Kim. 2023. Application of ensemble neural-network method to integrated sugar content prediction model for citrus fruit using Vis/NIR spectroscopy. *Journal of Food Engineering* 338 (2023), 111254.
- [30] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. 2023. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 4015–4026.
- [31] Manuel Knott, Fernando Perez-Cruz, and Thijs Defraeye. 2023. Facilitated machine learning for image-based fruit quality assessment. *Journal of Food Engineering* 345 (2023), 111401.
- [32] Jeroen Lammertyn, Ann Peirs, Josse De Baerdemaeker, and Bart Nicolai. 2000. Light penetration properties of NIR radiation in fruit with respect to non-destructive quality assessment. *Postharvest biology and technology* 18, 2 (2000), 121–132.
- [33] Jiaojiao Li, Chaoxiong Wu, Rui Song, Yunsong Li, and Fei Liu. 2020. Adaptive weighted attention network with camera spectral sensitivity prior for spectral reconstruction from RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 462–463.
- [34] Michael Li, Santoso Wibowo, Wei Li, and Lily D Li. 2021. Quantitative spectral data analysis using extreme learning machines algorithm incorporated with PCA. *Algorithms* 14, 1 (2021), 18.
- [35] Yuqi Li, Chong Wang, and Jieyu Zhao. 2017. Locally linear embedded sparse coding for spectral reconstruction from RGB images. *IEEE Signal Processing Letters* 25, 3 (2017), 363–367.
- [36] Xiaofeng Liu, Binghao Liu, Dong Jiang, Shiping Zhu, Wanxia Shen, Xin Yu, Yang Xue, Mengyu Liu, Jingyin Feng, and Xiaochun Zhao. 2019. The accumulation and composition of essential oil in kumquat peel. *Scientia Horticulturae* 252 (2019), 121–129.
- [37] Yutong Liu, Landu Jiang, Linghe Kong, Qiao Xiang, Xue Liu, and Guihai Chen. 2021. Wi-Fruit: See through fruits with smart devices. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 4 (2021), 1–29.

- [38] Delia Lorente, Nuria Aleixos, JUAN Gómez-Sanchis, Sergio Cubero, Oscar Leonardo García-Navarrete, and José Blasco. 2012. Recent advances and applications of hyperspectral imaging for fruit and vegetable quality assessment. *Food and Bioprocess Technology* 5 (2012), 1121–1142.
- [39] Xiankai Lu, Chao Ma, Bingbing Ni, Xiaokang Yang, Ian Reid, and Ming-Hsuan Yang. 2018. Deep regression tracking with shrinkage loss. In *Proceedings of the European conference on computer vision (ECCV)*. 353–369.
- [40] Lembe Samukelo Magwaza and Umezuruike Linus Opara. 2015. Analytical methods for determination of sugars and sweetness of horticultural products—A review. *Scientia Horticulturae* 184 (2015), 179–192.
- [41] Display Mate. [n. d.]. Light Spectra for the Apple iPhone X. [https://www.displaymate.com/Spectra\\_41a.html](https://www.displaymate.com/Spectra_41a.html).
- [42] Andrew JS McGonigle, Thomas C Wilkes, Tom D Perring, Jon R Willmott, Joseph M Cook, Forrest M Mims III, and Alfio V Parisi. 2018. Smartphone spectrometers. *Sensors* 18, 1 (2018), 223.
- [43] Anna Grazia Mignani, Andrea Azelio Mencaglia, Massimo Baldi, and Leonardo Ciaccheri. 2015. SpiderSpec: A low-cost compact colorimeter with IoT functionality. In *Fifth Asia-Pacific Optical Sensors Conference*, Vol. 9655. SPIE, 96–99.
- [44] mileseeey. 2023. SM20 Saccharimeter. <https://www.mileseeey.net/ProductCenter/ProductCenterList1034/10.html>.
- [45] Kumi Miyamoto and Yoshinobu Kitano. 1995. Non-destructive determination of sugar content in satsuma mandarin fruit by near infrared transmittance spectroscopy. *Journal of Near Infrared Spectroscopy* 3, 4 (1995), 227–237.
- [46] Carlos A Mora, Antoine F Herzog, Renzo A Raso, and Wendelin J Stark. 2015. Programmable living material containing reporter micro-organisms permits quantitative detection of oligosaccharides. *Biomaterials* 61 (2015), 1–9.
- [47] National Institute of Health. 2020. Healthcare Costs of Poor Nutrition. <https://www.nih.gov/news-events/news-releases/healthcare-costs-poor-nutrition> Accessed: 2023-10-01.
- [48] P Pathmanaban, BK Gnanavel, and Shanmuga Sundaram Anandan. 2019. Recent application of imaging techniques for fruit quality assessment. *Trends in Food Science & Technology* 94 (2019), 32–42.
- [49] Artzai Picon, Arantza Bereciartua-Perez, Itziar Eguskiza, Javier Romero-Rodriguez, Carlos Javier Jimenez-Ruiz, Till Eggers, Christian Klukas, and Ramon Navarra-Mestre. 2022. Deep convolutional neural network for damaged vegetation segmentation from RGB images based on virtual NIR-channel estimation. *Artificial Intelligence in Agriculture* 6 (2022), 199–210.
- [50] Susan L Pollack. 2001. Consumer demand for fruit and vegetables: the US example. *Changing structure of global food consumption and trade* 6 (2001), 49–54.
- [51] Evia Zunita D Pratiwi, Muhammad FR Pahlawan, Diah N Rahmi, Hanim Z Amanah, and Rudiati E Masithoh. 2023. Non-destructive evaluation of soluble solid content in fruits with various skin thicknesses using visible–shortwave near-infrared spectroscopy. *Open Agriculture* 8, 1 (2023), 20220183.
- [52] Aifeng Ren, Adnan Zahid, Ahmed Zoha, Syed Aziz Shah, Muhammad Ali Imran, Akram Alomainy, and Qammer H Abbasi. 2019. Machine learning driven approach towards the quality assessment of fresh fruits using non-invasive sensing. *IEEE Sensors Journal* 20, 4 (2019), 2075–2083.
- [53] Harish Sasikumar, Vishnu Prasad, Parama Pal, and Manoj M Varma. 2016. Diffractive interference optical analyzer (DiOPTER). In *Optical Diagnostics and Sensing XVI: Toward Point-of-Care Diagnostics*, Vol. 9715. SPIE, 18–21.
- [54] SCiO. [n. d.]. SCiO Mini. <https://shop.consumerphysics.com/>.
- [55] Neha Sharma, Muhammad Shahzaib Waseem, Shahrzad Mirzaei, and Mohamed Hefeeda. 2023. MobiSpectral: Hyperspectral Imaging on Mobile Devices. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*. 1–15.
- [56] Zhan Shi, Chang Chen, Zhiwei Xiong, Dong Liu, and Feng Wu. 2018. Hscnn+: Advanced cnn-based hyperspectral recovery from rgb images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 939–947.
- [57] Weiran Song, Nanfeng Jiang, Hui Wang, and Jordan Vincent. 2020. Use of smartphone videos and pattern recognition for food authentication. *Sensors and Actuators B: Chemical* 304 (2020), 127247.
- [58] Weiran Song, Yong-Huan Yun, Hui Wang, Zongyu Hou, and Zhe Wang. 2021. Smartphone detection of minced beef adulteration. *Microchemical Journal* 164 (2021), 106088.
- [59] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. 2014. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research* 15, 1 (2014), 1929–1958.
- [60] Sarun Sumriddetchkajorn, Kosom Chaitavon, and Yuttana Intaravanne. 2014. Mobile-platform based colorimeter for monitoring chlorine concentration in water. *Sensors and Actuators B: Chemical* 191 (2014), 561–566.
- [61] Donald F Swinehart. 1962. The beer-lambert law. *Journal of chemical education* 39, 7 (1962), 333.
- [62] Sandeep Kumar Vashist, Thomas van Oordt, E Marion Schneider, Roland Zengerle, Felix von Stetten, and John HT Luong. 2015. A smartphone-based colorimetric reader for bioanalytical applications using the screen-based bottom illumination provided by gadgets. *Biosensors and Bioelectronics* 67 (2015), 248–255.
- [63] Maurizio Ventura, Anton de Jager, Herman de Putter, and Frans PMM Roelofs. 1998. Non-destructive determination of soluble solids in apple fruit by near infrared spectroscopy (NIRS). *Postharvest Biology and Technology* 14, 1 (1998), 21–27.
- [64] Dhruv Verma, Ian Ruffolo, David B Lindell, Kiriakos N Kutulakos, and Alex Mariakakis. 2024. ChromaFlash: Snapshot Hyperspectral Imaging Using Rolling Shutter Cameras. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 8, 3 (2024), 1–31.

- [65] Wikipedia. 2024. CIE 1931 color space. [https://en.wikipedia.org/wiki/CIE\\_1931\\_color\\_space](https://en.wikipedia.org/wiki/CIE_1931_color_space).
- [66] Zhiwei Xiong, Zhan Shi, Huiqun Li, Lizhi Wang, Dong Liu, and Feng Wu. 2017. Hscnn: Cnn-based hyperspectral image recovery from spectrally undersampled projections. In *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 518–525.
- [67] Yiqi Yan, Lei Zhang, Jun Li, Wei Wei, and Yanning Zhang. 2018. Accurate spectral super-resolution from single RGB image using multi-scale CNN. In *Pattern Recognition and Computer Vision: First Chinese Conference, PRCV 2018, Guangzhou, China, November 23-26, 2018, Proceedings, Part II 1*. Springer, 206–217.
- [68] Xin Yang, Yanru Wang, Wei Liu, Yuhuan Zhang, Fangqing Zheng, Shuaixing Wang, Daohong Zhang, and Jianlong Wang. 2016. A portable system for on-site quantification of formaldehyde in air based on G-quadruplex halves coupled with A smartphone reader. *Biosensors and Bioelectronics* 75 (2016), 48–54.
- [69] Zhiyang Yao, Shuyang Liu, Xiaoyun Yuan, and Lu Fang. 2024. SPECAT: SPatial-spEctral Cumulative-Attention Transformer for High-Resolution Hyperspectral Image Reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 25368–25377.
- [70] Hui-Shyong Yeo, Juyoung Lee, Andrea Bianchi, David Harris-Birtill, and Aaron Quigley. 2017. Specam: Sensing surface color and material with the front-facing camera of a mobile device. In *Proceedings of the 19th International Conference on Human-Computer Interaction with Mobile Devices and Services*. 1–9.
- [71] Fikret YILDIZ, Ahmet Turan ÖZDEMİR, and Selman Uluişik. 2018. Custom design fruit quality evaluation system with non-destructive testing (NDT) techniques. In *2018 International Conference on Artificial Intelligence and Data Processing (IDAP)*. IEEE, 1–5.
- [72] Baohua Zhang, Dejian Dai, Jichao Huang, Jun Zhou, Qifa Gui, and Fang Dai. 2018. Influence of physical and biological variability and solution methods in fruit and vegetable quality nondestructive inspection by using imaging and near-infrared spectroscopy techniques: A review. *Critical reviews in food science and nutrition* 58, 12 (2018), 2099–2118.
- [73] Jiakun Zhang, Liu Zhang, Ying Song, and Yan Zheng. 2023. The algorithm research of low-rank matrix spectral reconstruction for ground targets. *Results in Physics* 52 (2023), 106868.
- [74] Xin-xin Zhang, Pao Li, Mei Yu, Li-wen Jiang, Xia Liu, and Yang Shan. 2022. Progress in non-destructive citrus quality detection using near-infrared spectroscopy. (2022).
- [75] Yuzhi Zhao, Lai-Man Po, Qiong Yan, Wei Liu, and Tingyu Lin. 2020. Hierarchical regression network for spectral reconstruction from RGB images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*. 422–423.