

Apéndice B. Procedimientos complementarios de preprocesamiento

Una vez adquiridas y organizadas las imágenes RGB y los cubos espectrales descritas en el [Error! No se encuentra el origen de la referencia.](#), se realizaron diferentes procedimientos de preprocesamiento con el propósito de preparar la información para las etapas de entrenamiento, validación e inferencia de los modelos de inteligencia artificial implementados. Estas operaciones incluyeron procesos de estandarización de las imágenes RGB, corrección y análisis de la información espectral, segmentación automática de las muestras y construcción de las representaciones empleadas por los modelos de detección y clasificación.

Adicionalmente, se definieron las estructuras de almacenamiento correspondientes a los conjuntos de datos generados, permitiendo mantener la trazabilidad de las muestras y facilitar su utilización durante las diferentes etapas del desarrollo del sistema.

B.1 Redimensionamiento de imágenes RGB

Con el propósito de estandarizar las dimensiones de las muestras y adaptarlas a los requerimientos de entrada de la red neuronal convolucional, las imágenes RGB fueron sometidas a un proceso de redimensionamiento. Durante la etapa de adquisición y almacenamiento descrita en el [Error! No se encuentra el origen de la referencia.](#), las imágenes conservaron su resolución original; posteriormente, durante el preprocesamiento, todas las muestras fueron escaladas a una resolución de 100×100 píxeles.

Este procedimiento fue implementado mediante scripts desarrollados en Python utilizando TensorFlow, donde las imágenes fueron cargadas automáticamente desde las carpetas correspondientes a cada clase y redimensionadas durante la construcción del conjunto de datos empleado en el entrenamiento del modelo.

Código B.1. *Fragmento del script utilizado para el redimensionamiento automático de las imágenes RGB.*

```
1 IMG_SIZE = (100, 100)
2
3 dataset = tf.keras.utils.image_dataset_from_directory(
4     DATASET_PATH,
5     labels='inferred',
6     label_mode='int',
7     class_names=['Buenas', 'Malas'],
8     image_size=IMG_SIZE,
9     batch_size=BATCH_SIZE,
10    shuffle=True,
11    seed=42
12 )
```

Como resultado de este proceso, se obtuvieron imágenes con dimensiones uniformes de 100×100 píxeles como se muestra en la figura B.1, manteniendo las características visuales relevantes del fruto y reduciendo la complejidad computacional durante las etapas de entrenamiento y validación. La estandarización de las muestras permitió garantizar entradas homogéneas para la red neuronal convolucional y facilitó el procesamiento del conjunto de datos.

Figura B.1 *Ejemplo de redimensionamiento aplicado a las imágenes RGB.*



B.2 Determinación de la referencia espectral

Debido a las variaciones observadas en la respuesta espectral entre las diferentes sesiones de adquisición, se implementó un procedimiento para determinar una referencia espectral individual para cada cubo espectral. Estas variaciones se encontraban asociadas principalmente a pequeños desplazamientos en la configuración del sistema óptico, los cuales producían cambios en la ubicación de la región de mayor intensidad dentro del eje espectral.

Por esta razón, en lugar de utilizar una banda de referencia fija para todas las capturas, se determinó individualmente la banda de máxima intensidad media de cada cubo. Esta banda, denominada λ_{pico} , fue utilizada posteriormente como referencia para seleccionar las cinco bandas espectrales empleadas durante la construcción de los conjuntos de datos.

B.2.1 Obtención del perfil espectral promedio

Los cubos espectrales adquiridos mediante la cámara SWIR fueron almacenados mediante una estructura tridimensional C , compuesta por un eje espacial, un eje espectral y un eje temporal.

En la configuración empleada, el eje espacial estuvo conformado por 1024 posiciones, el eje espectral por 1280 bandas y el número de líneas temporales dependió de la duración de cada adquisición. La extracción de una imagen bidimensional correspondiente a una banda específica b se realizó fijando la posición del eje espectral del cubo.

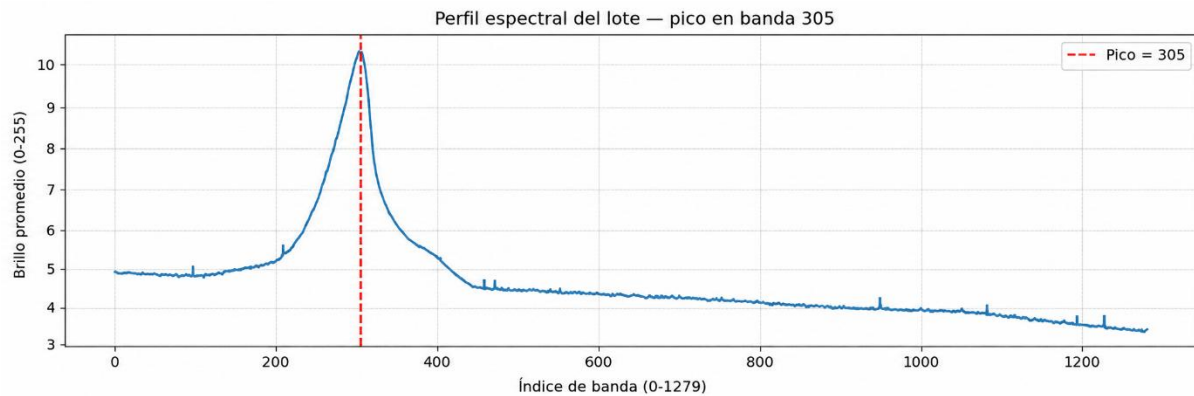
Para caracterizar la distribución global de intensidad de cada captura, se calculó el perfil espectral promedio mediante la media de los valores registrados sobre los ejes espacial y temporal:

$$\bar{I}(\lambda) = \frac{1}{Y \cdot T} \sum_{y=1}^Y \sum_{t=1}^T C(y, \lambda, t) \quad (\text{Ec. b.1})$$

donde $C(y, \lambda, t)$ representa la intensidad registrada en la posición espacial y , la banda espectral λ y el instante temporal t ; mientras que Y y T corresponden al número de posiciones espaciales y líneas temporales del cubo, respectivamente.

Como resultado de este procedimiento se obtuvo un perfil unidimensional representativo de la variación de la intensidad media a lo largo del eje espectral. Este perfil permitió identificar la región de máxima respuesta registrada por el sistema de adquisición para cada cubo procesado.

Figura B.2. Perfil espectral promedio obtenido a partir de un cubo espectral.



B.2.2 Determinación de la banda pico

A partir del perfil espectral promedio se determinó automáticamente el índice correspondiente a la banda de mayor intensidad media. Esta posición fue definida como la banda pico de cada cubo y se obtuvo mediante:

$$\lambda_{pico} = \arg \max_{\lambda} [I(\lambda)] \quad (\text{Ec. b.2})$$

El cálculo se realizó directamente sobre el cubo original antes de aplicar los procedimientos posteriores de procesamiento. La determinación individual de λ_{pico} permitió compensar los desplazamientos observados en la respuesta del sistema óptico entre las diferentes sesiones de adquisición.

De esta manera, las bandas utilizadas posteriormente no fueron establecidas mediante posiciones absolutas dentro del eje espectral, sino mediante desplazamientos relativos respecto a la banda pico de cada cubo. Esta estrategia permitió mantener una referencia consistente durante el procesamiento de capturas obtenidas en diferentes sesiones experimentales.

B.2.3 Generación del reporte de picos espectrales

La determinación de la banda pico se realizó previamente a la construcción de los conjuntos de datos. Para ello, se implementó un procedimiento automatizado encargado de recorrer los cubos espectrales disponibles, calcular el perfil espectral promedio e identificar la posición de máxima intensidad correspondiente a cada captura.

Los resultados fueron almacenados en el archivo **reporte_picos.json**, conservando la relación entre cada cubo procesado y su respectivo valor de λ_{pico} . Este procedimiento evitó

recalcular la referencia espectral durante cada ejecución del pipeline de generación de los conjuntos de datos y permitió mantener la trazabilidad de las bandas utilizadas para cada muestra.

Una vez determinada la referencia individual de cada cubo, las cinco bandas utilizadas por los modelos fueron seleccionadas mediante λ_{pico} , Este procedimiento se describe en la siguiente sección.

B.3 Selección de bandas espectrales

Una vez determinada la banda de referencia λ_{pico} para cada cubo espectral, se realizó la selección de las bandas utilizadas durante la construcción de las entradas de los modelos de detección y clasificación. Debido a los desplazamientos observados en la respuesta espectral entre las diferentes sesiones de adquisición, las bandas no fueron seleccionadas mediante índices absolutos, sino a partir de desplazamientos relativos respecto a la banda pico de cada cubo.

De esta manera, para cada banda seleccionada se estableció un desplazamiento δ_i , obteniendo su posición dentro del cubo mediante:

$$\lambda_i = \lambda_{pico} + \delta_i \quad (\text{Ec. b.3})$$

donde λ_i corresponde al índice de la banda seleccionada, λ_{pico} representa la banda de máxima intensidad media determinada previamente y δ_i corresponde al desplazamiento aplicado respecto a dicha referencia.

La configuración final estuvo compuesta por cinco bandas espectrales, cuyos desplazamientos relativos fueron $[+11, +2, -5, -10, -3]$. Las tres primeras bandas fueron utilizadas para construir las representaciones pseudo-RGB destinadas al modelo de detección, mientras que

las cinco bandas conformaron los tensores espectrales empleados como entrada del modelo de clasificación.

B.3.1 Selección mediante offsets relativos

El uso desplazamientos relativos permitió adaptar automáticamente la selección de bandas a las características espectrales de cada cubo adquirido. Para ello, una vez recuperado el valor de λ_{pico} almacenado durante la etapa anterior, se calcularon las posiciones correspondientes a cada uno de los cinco canales mediante:

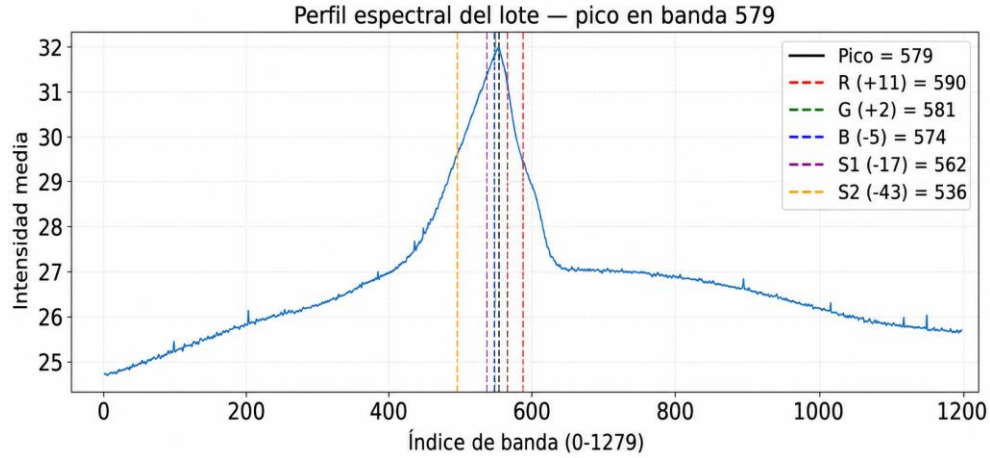
$$\delta = [+11, +2, -5, -10, -3] \text{ (Ec. b.4)}$$

De esta manera, la configuración espectral utilizada para cada cubo quedó definida como:

$$\lambda = [\lambda_{pico} + 11, \lambda_{pico} + 2, \lambda_{pico} - 5, \lambda_{pico} - 10, \lambda_{pico} - 3] \text{ (Ec. b.5)}$$

Este procedimiento permitió conservar la posición relativa de las bandas seleccionadas respecto a la región de máxima respuesta del sistema, reduciendo el efecto de los desplazamientos observados entre diferentes sesiones de adquisición.

Figura B.3. *Distribución de las bandas espectrales seleccionadas mediante desplazamientos relativos respecto a la banda de máxima intensidad media.*



B.3.2 Configuración espectral de cinco bandas

A partir de los desplazamientos establecidos respecto a λ_{pico} , se definió una configuración compuesta por cinco bandas espectrales. Esta configuración fue utilizada de manera uniforme durante la construcción de las muestras destinadas a los modelos de detección y clasificación.

La configuración empleada se resume en la Tabla B.1.

Tabla B.1. Configuración de las bandas espectrales utilizadas para la construcción de los conjuntos de datos.

Canal	Posición relativa
1	$\lambda_{pico} + 11$
2	$\lambda_{pico} + 2$
3	$\lambda_{pico} - 5$
4	$\lambda_{pico} - 10$

La utilización de esta configuración permitió mantener una estructura espectral consistente entre las diferentes muestras procesadas, tomando como referencia la posición individual de λ_{pico} determinada para cada cubo.

B.3.3 Bandas destinadas al pseudo-RGB y bandas espectrales adicionales

Las cinco bandas seleccionadas fueron distribuidas de acuerdo con los requerimientos de entrada de los modelos implementados. Las bandas ubicadas en los desplazamientos +11, +2 y -5 respecto a λ_{pico} fueron destinadas a la construcción de las representaciones pseudo-RGB utilizadas por el modelo de detección YOLO.

Por otra parte, las bandas correspondientes a los desplazamientos -10 y -3 fueron conservadas como canales espectrales adicionales. De esta manera, las muestras destinadas al modelo de clasificación quedaron constituidas por tensores de cinco canales, integrando las tres bandas empleadas para la representación pseudo-RGB y las dos bandas espectrales complementarias.

Esta organización permitió obtener, a partir de una misma configuración espectral, las estructuras de entrada requeridas por ambos modelos. Mientras el modelo de detección utilizó representaciones pseudo-RGB de tres canales, el modelo de clasificación empleó la información contenida en las cinco bandas seleccionadas. Los procedimientos posteriores de segmentación automática y extracción de las regiones de interés se describen en la siguiente sección.

B.4 Segmentación automática y extracción de regiones de interés

Una vez seleccionadas las bandas espectrales de interés, se implementó un procedimiento de segmentación automática para identificar las mandarinas presentes en cada cubo y extraer las

regiones correspondientes a cada fruto. Este proceso permitió transformar las escenas espectrales originales, que podían contener múltiples mandarinas, en muestras individuales destinadas a la posterior construcción de los conjuntos de datos.

La segmentación se realizó utilizando una imagen bidimensional de referencia obtenida a partir de la información espectral del cubo. Sobre esta imagen se aplicaron procedimientos de normalización, umbralización automática y operaciones morfológicas para generar una máscara binaria que permitiera diferenciar los frutos del fondo. Seguido de esto, se identificaron los contornos presentes en la máscara, se calcularon sus cajas delimitadoras y se extrajeron las regiones de interés correspondientes.

B.4.1 Normalización de la imagen de referencia

Para realizar la segmentación de las mandarinas se utilizó como referencia la banda de máxima intensidad media determinada previamente para cada cubo. La imagen bidimensional correspondiente a esta banda fue extraída mediante:

$$I_{ref} = C[:, \lambda_{pico}, :] \text{ (Ec. b.6)}$$

Donde C representa el cubo espectral y corresponde al índice de la banda de referencia. Debido a la variabilidad en la intensidad entre las diferentes capturas, la imagen fue normalizada en el intervalo $[0, 255]$ antes de aplicar el procedimiento de segmentación:

$$I_N = \frac{I_{ref} - \min(I_{ref})}{\max(I_{ref}) - \min(I_{ref})} \cdot 255 \text{ (Ec. b.7)}$$

Una vez normalizada la imagen fue convertida a formato uint8, donde se permitió aplicar los procedimientos posteriores de procesamiento de imágenes utilizados para la generación de la máscara binaria.

B.4.2 Umbralización mediante Otsu

Una vez normalizada la imagen de referencia, se aplicó el método de umbralización automática de Otsu para separar las regiones correspondientes a las mandarinas del fondo de la escena. Este procedimiento determina automáticamente un valor de umbral t^* mediante la maximización de la varianza entre las clases resultantes:

$$t^* = \arg \max_t \sigma_B^2(t) \quad (\text{Ec. b.8})$$

Donde $\sigma_B^2(t)$ es la varianza entre las clases generadas para un determinado valor de umbral t . A partir de este umbral se generó una máscara binaria M , definida como:

$$M(y, t) = \begin{cases} 1, & I_N(y, t) > t^* \\ 0, & I_N(y, t) \leq t^* \end{cases} \quad (\text{Ec. b.9})$$

B.4.3 Operaciones morfológicas

La máscara obtenida mediante la umbralización podía contener pequeñas regiones aisladas, discontinuidades y otras irregularidades asociadas al ruido presente en las imágenes. Por esta razón, se aplicaron operaciones morfológicas destinadas a mejorar la estructura de las regiones segmentadas antes de realizar la identificación de los frutos.

El procedimiento implementado consistió en una operación de cierre morfológico seguida de una operación de apertura. El cierre permitió rellenar pequeñas discontinuidades presentes dentro de las regiones correspondientes a las mandarinas, mientras que la apertura facilitó la eliminación de regiones aisladas de menor tamaño.

Estas operaciones pueden expresarse como:

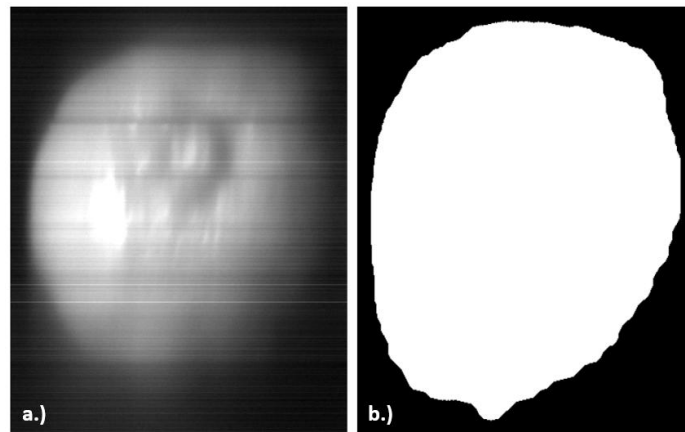
$$M_c = (M \oplus K) \ominus K \quad (\text{Ec. b.10})$$

$$M_a = (M_c \ominus K) \oplus K \quad (\text{Ec. b.11})$$

donde M representa la máscara binaria original, M_c la máscara obtenida después del cierre, M_a la máscara resultante después de la apertura y K el elemento estructurante empleado durante las operaciones de dilatación ($\oplus$) y erosión ($\ominus$).

Como resultado, se obtuvo una máscara binaria con regiones más continuas y una menor presencia de elementos aislados, facilitando la identificación posterior de los contornos asociados a cada fruto.

Figura B.4. *Etapas del procedimiento de segmentación automática: a) imagen gray scale de una mandarina luego del recorte, b) máscara resultante después de las operaciones morfológicas.*



B.4.4 Extracción de bounding boxes

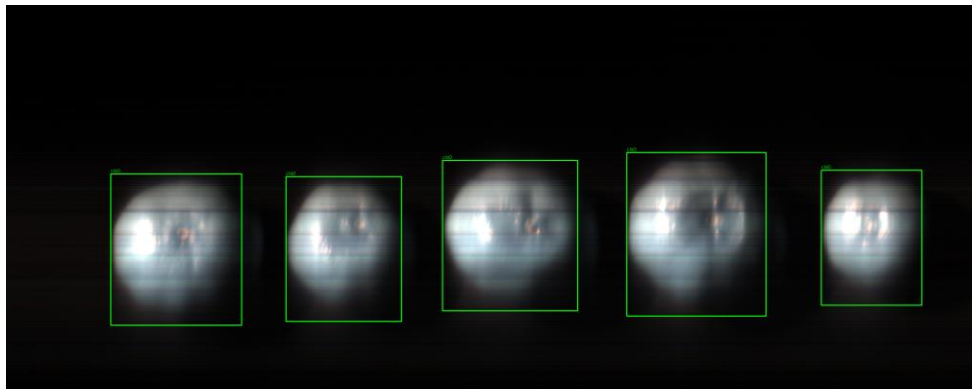
A partir de la máscara binaria procesada se realizó la detección de los contornos correspondientes a las mandarinas presentes en la escena. Para cada contorno identificado se calculó una caja delimitadora o bounding box, definida mediante:

$$B_i = (x_i, y_i, w_i, h_i) \quad (\text{Ec. b.12})$$

Donde x_i y y_i corresponden a las coordenadas de la esquina superior izquierda de la caja delimitadora, mientras que, w_i y h_i representan su ancho y altura, respectivamente.

Con el propósito de reducir la detección de regiones asociadas al ruido residual, se aplicó un criterio de área mínima para descartar contornos de tamaño reducido. De esta manera, únicamente las regiones que cumplieran las condiciones establecidas fueron consideradas como mandarinas válidas y utilizadas durante la extracción de las muestras individuales.

Figura B.5. *Identificación automática de las mandarinas presentes en el cubo espectral mediante cajas delimitadoras.*



B.4.5 Generación de regiones de interés

Finalmente, las coordenadas de las cajas delimitadoras fueron utilizadas para extraer directamente del cubo original las regiones de interés correspondientes a cada mandarina detectada. Para una caja delimitadora $B_i = (x_i, y_i, w_i, h_i)$ la región de interés puede expresarse mediante:

$$ROI_i = C[y_i : y_i + h_i, : , x_i : x_i + w_i] \quad (\text{Ec. b.13})$$

De esta manera, la extracción se realizó sobre el cubo espectral original y no sobre la imagen normalizada utilizada durante la segmentación, permitiendo conservar la información espectral de las bandas seleccionadas para cada fruto.

Las regiones obtenidas fueron utilizadas posteriormente para generar las muestras individuales destinadas al modelo de clasificación. En el caso del conjunto de datos utilizado por YOLO, las cajas delimitadoras fueron empleadas para generar las anotaciones correspondientes a la ubicación de los frutos dentro de las escenas pseudo-RGB.

Este procedimiento permitió automatizar la identificación y extracción de múltiples mandarinas presentes dentro de un mismo cubo espectral, manteniendo la correspondencia entre la información espacial de cada fruto y sus características espectrales.

B.5 Construcción de los conjuntos de datos

Una vez seleccionadas las bandas espectrales y obtenidas las regiones correspondientes a las mandarinas presentes en cada cubo, se procedió a construir las representaciones utilizadas como entrada de los modelos de detección y clasificación. Debido a las diferencias existentes entre los requerimientos de ambos modelos, se implementaron procedimientos independientes para la generación de las imágenes pseudo-RGB y de los tensores espectrales de cinco canales.

Para el modelo de detección, las tres bandas destinadas a la representación pseudo-RGB fueron procesadas mediante normalización independiente y corrección gamma antes de conformar la imagen final. Por otra parte, para el modelo de clasificación se conservaron las cinco bandas seleccionadas en formato float32, manteniendo la información espectral de cada región de interés para su posterior procesamiento durante el entrenamiento.

B.5.1 Normalización por canal

Las intensidades registradas en las bandas espectrales podían presentar diferentes rangos de valores, por lo que cada una de las tres bandas destinadas a la construcción del pseudo-RGB fue normalizada de manera independiente. Para una banda espectral I_b , la normalización se realizó mediante:

$$I_{N,b} = \frac{I_b - \min(I_b)}{\max(I_b) - \min(I_b)} \text{ (Ec. b.14)}$$

Donde I_b representa la imagen correspondiente a la banda espectral seleccionada e $I_{N,b}$ la imagen normalizada en el intervalo $[0,1]$.

La normalización independiente permitió ajustar el rango dinámico de cada canal antes de realizar su combinación, evitando que las diferencias de intensidad entre las bandas dominaran la representación visual resultante.

B.5.2 Corrección gamma

Después de la normalización, se aplicó una transformación gamma sobre cada uno de los canales destinados a la generación del pseudo-RGB. Esta operación permitió modificar la distribución de las intensidades y mejorar la representación de las estructuras presentes en las imágenes espectrales.

$$I_{\gamma,b} = I_{N,b}^{\gamma} \text{ (Ec. b.15)}$$

Donde $I_{N,b}$ corresponde a la banda normalizada, γ al factor de corrección empleado e $I_{\gamma,b}$ a la imagen resultante. Una vez aplicada la transformación, los valores fueron escalados al intervalo $[0,255]$ y convertidos al formato uint8, requerido para el almacenamiento de las imágenes utilizadas durante el entrenamiento del modelo de detección.

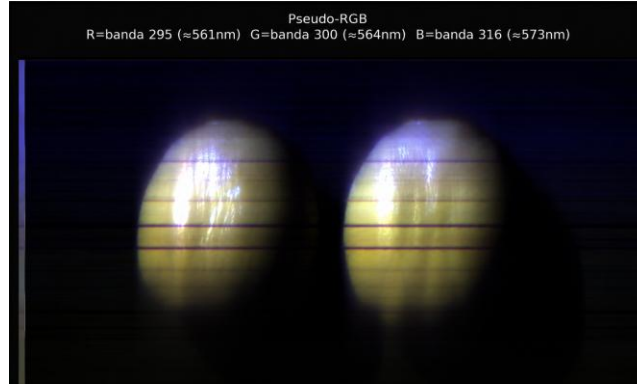
B.5.3 Generación de imágenes pseudo-RGB

Las representaciones pseudo-RGB fueron construidas mediante la combinación de las tres bandas espectrales ubicadas en las posiciones relativas $\lambda_{pico} + 11$, $\lambda_{pico} + 2$, $\lambda_{pico} - 5$. Después de aplicar los procedimientos de normalización independiente y corrección gamma, las imágenes resultantes fueron agrupadas para conformar una estructura tridimensional de tres canales:

$$I_{pseudo-RGB} = stack(I_{\gamma,+11}, I_{\gamma,+2}, I_{\gamma,-5}) \text{ (Ec. b.16)}$$

Las imágenes obtenidas conservaron una estructura de tres canales compatible con el modelo YOLO, aunque sus componentes no correspondían a los canales rojo, verde y azul de una imagen convencional, sino a información procedente de diferentes regiones del espectro adquiridas mediante la cámara SWIR.

Figura B.6. Construcción de la representación pseudo-RGB a partir de tres bandas espectrales seleccionadas respecto a la banda pico.



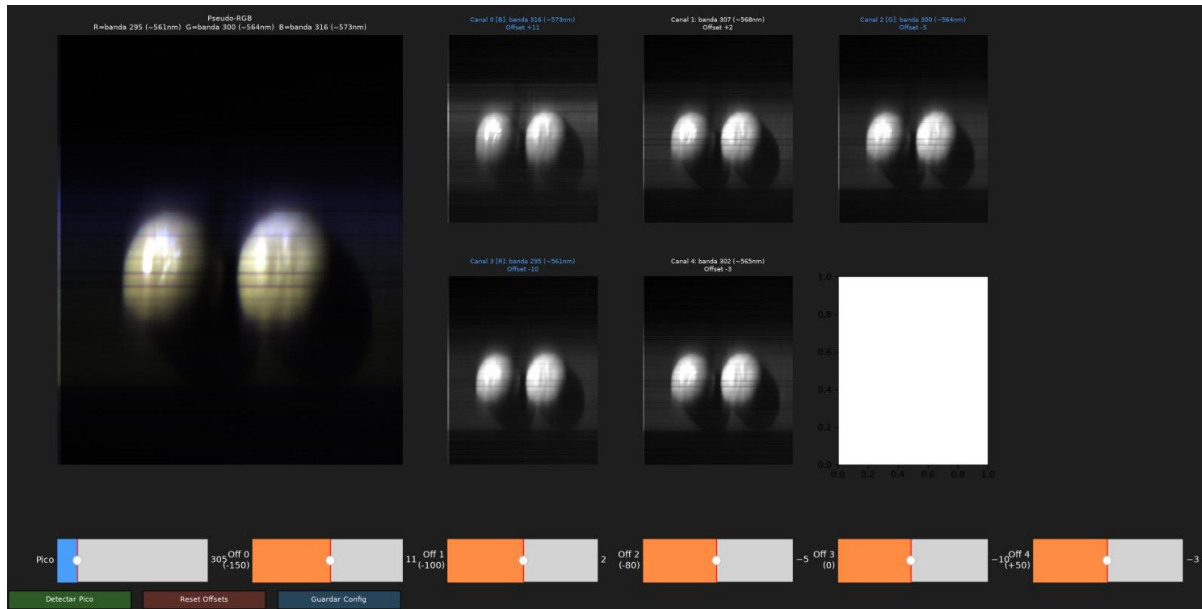
B.5.4 Construcción de tensores espectrales de cinco bandas

Para la construcción de las muestras destinadas al modelo de clasificación se utilizaron las regiones de interés extraídas previamente del cubo espectral. Para cada mandarina se recuperó la información correspondiente a las cinco bandas seleccionadas y se generó un tensor espectral mediante:

$$X_i = \text{stack}(I_{i,+11}, I_{i,+2}, I_{i,-5}, I_{i,-10}, I_{i,-3}) \text{ (Ec. b.17)}$$

Donde es la representación de tensor que corresponde a la muestra y cada componente corresponde a la región de interés extraída en la posición. A diferencia de las representaciones pseudo-RGB, los tensores espectrales fueron almacenados en formato float32 sin aplicar durante esta etapa la normalización utilizada para la generación de las imágenes de tres canales. Esta decisión permitió conservar los valores espectrales de las muestras y aplicar posteriormente los procedimientos de normalización requeridos durante el entrenamiento del modelo de clasificación.

Figura B.7. Representación de las cinco bandas espectrales que conforman el tensor espectral correspondiente a una región de interés.



B.6 Organización y almacenamiento de los conjuntos de datos

Una vez generadas las representaciones espectrales destinadas a los modelos de detección y clasificación, se realizó la organización, depuración y almacenamiento de las muestras obtenidas. Debido a las diferencias entre las estructuras de entrada requeridas por ambos modelos, se construyeron conjuntos de datos independientes, conservando en cada caso la información necesaria para garantizar la trazabilidad entre las muestras procesadas y los cubos espectrales de origen.

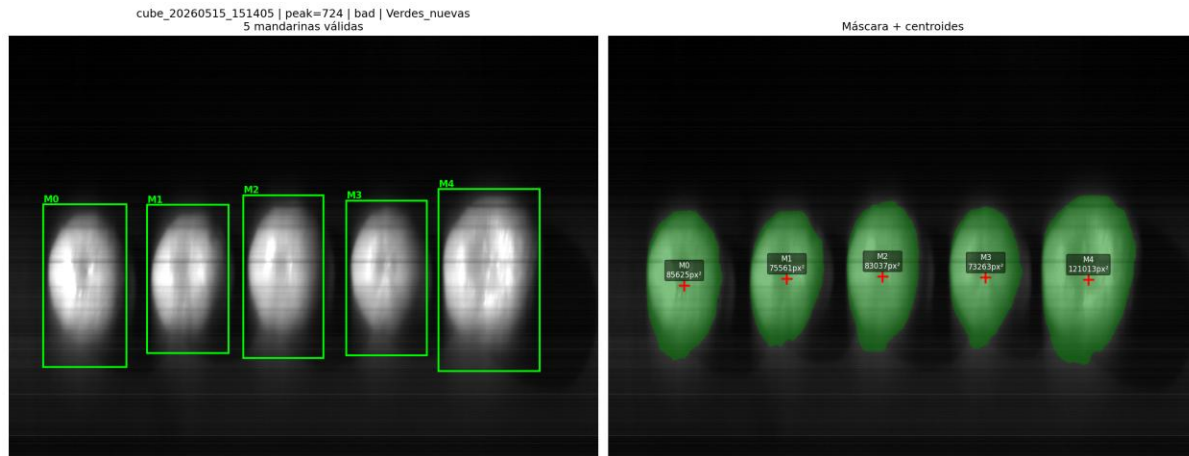
B.6.1 Dataset para el modelo CNN

El dataset para el modelo de clasificación se construyó a partir de las regiones de interés extraídas individualmente para cada mandarina. Cada muestra fue almacenada como un tensor espectral compuesto por las cinco bandas seleccionadas, utilizando formato float32.

Los datos fueron organizados de acuerdo con su categoría correspondiente a frutos aptos y no aptos. Adicionalmente se conservó información del cubo espectral de origen, esto con el fin de mantener la trazabilidad de cada región de interés durante las etapas posteriores de entrenamiento y evaluación.

Una vez generado el conjunto de datos, se realizó una inspección de las muestras obtenidas con el propósito de identificar regiones incorrectamente segmentadas, recortes incompletos y otros casos que no representaran adecuadamente los frutos capturados. Las muestras identificadas durante este procedimiento fueron excluidas antes de establecer la estructura definitiva del conjunto de datos. Como resultado, se obtuvo un conjunto de 694 muestras espectrales válidas para su posterior etapa de entrenamiento y validación.

Figura B.8. *Ejemplos de regiones de interés espectrales pertenecientes a las categorías apta y no apta después del proceso de depuración.*



Para evitar la transferencia de información entre los conjuntos utilizados durante el desarrollo y evaluación del modelo, la separación de las muestras se realizó considerando el cubo espectral de origen. De esta manera, todas las regiones de interés extraídas de un mismo cubo fueron asignadas exclusivamente a una única partición.

B.6.2 Dataset para YOLO

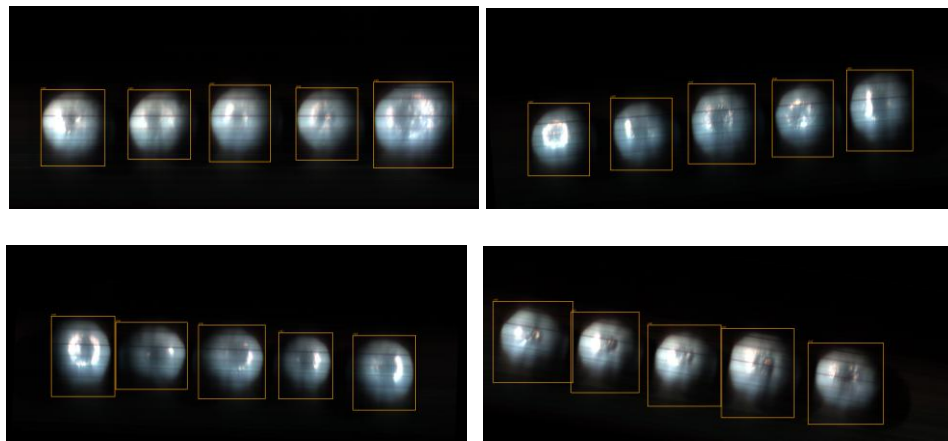
El conjunto de datos destinado al modelo de detección fue construido a partir de las representaciones pseudo-RGB correspondientes a las escenas completas. Para cada imagen generada se almacenó la representación de tres canales en formato uint8 junto con un archivo de anotación que contenía la información espacial de las mandarinas identificadas durante el procedimiento de segmentación automática.

Las muestras fueron sometidas posteriormente a una etapa de inspección y depuración con el propósito de eliminar imágenes o anotaciones que presentaran inconsistencias. Como resultado de este procedimiento se obtuvo un conjunto compuesto por 112 imágenes originales, las cuales fueron utilizadas para establecer las particiones destinadas al entrenamiento y validación del modelo de detección.

La división del conjunto de datos produjo 90 imágenes para entrenamiento y 22 imágenes para validación. Posteriormente, se aplicaron procedimientos de aumento de datos exclusivamente sobre el conjunto de entrenamiento, evitando generar muestras aumentadas a partir de las imágenes reservadas para validación.

Como resultado de este procedimiento se generaron 357 imágenes aumentadas a partir del conjunto de entrenamiento original obteniendo un total de 447 datos para entrenamiento. La aplicación del aumento de datos permitió incrementar la variabilidad de las muestras disponibles para el ajuste del modelo, manteniendo independiente el conjunto utilizado durante la evaluación.

Figura B.9. Ejemplos de representaciones pseudo-RGB y anotaciones utilizadas para la construcción del conjunto de datos destinado al modelo YOLO.



B.6.3 Generación de anotaciones YOLO

Las cajas delimitadoras obtenidas durante la etapa de segmentación automática fueron utilizadas para generar los archivos de anotación requeridos por el modelo YOLO, para cada mandarina identificada se transformaron las coordenadas de la caja delimitadora al formato normalizado utilizado por el modelo.

Para una imagen con dimensiones $W \times H$ una caja delimitadora definida por (x, y, w, h) , las coordenadas normalizadas del centro y las dimensiones se determinaron mediante:

$$x_c = \frac{x + \frac{w}{2}}{W} \quad (\text{Ec. b.18})$$

$$y_c = \frac{y + \frac{h}{2}}{H} \quad (\text{Ec. b.19})$$

$$w_n = \frac{w}{W}; \quad h_n = \frac{h}{H} \quad (\text{Ec. b.20})$$

La organización final permitió mantener la correspondencia entre las representaciones pseudo-RGB y sus respectivas anotaciones, facilitando su utilización durante el ajuste del modelo de detección.

Figura B.10 Estructura general de organización y almacenamiento de los conjuntos de datos destinados a los modelos de clasificación y detección.

