

1. Segmentación

Uno de los principales problemas de nuestro set de datos es que las imágenes de flores del azafrán no se encuentran etiquetadas y, por lo tanto, una de las tareas que será necesario llevar a cabo es la de etiquetar todo el dataset.

Este paso se puede llevar a cabo manualmente con herramientas como label-studio, que además son de código abierto. Sin embargo, la tarea es muy tediosa y, si el dataset es grande, puede llevar demasiado tiempo del proyecto.

Como alternativa, vamos a preparar dos algoritmos para segmentar de forma automática las imágenes, y en aquellas en las que, por la naturaleza de la propia imagen, no se consiga una buena segmentación, haremos el proceso manual.

Estos algoritmos están planteados desde dos enfoques. El primero será utilizando técnicas clásicas de visión artificial y el segundo estará basado en el paper de 2018 "Unsupervised image segmentation by backpropagation" de Asako Kanezaki. [Kanezaki, 2018], que utiliza redes convolucionales para extraer características de las imágenes.

1.1. Segmentación con Visión clásica

Para el algoritmo de visión clásica hemos utilizado watershed. Watershed es una técnica de segmentación basada en morfología matemática, que permite extraer las fronteras de las regiones que hay en una imagen, por eso también el proceso que vamos a realizar se puede catalogar como segmentación basada en regiones. Todos los pasos seguidos vienen representados en el esquema 1.1.

Partiremos de la imagen en formato RGB con tamaño (3024x4032)x3. Pasaremos a una imagen en escala de grises con tamaño (3024x4032)x1. A continuación aplicaremos dos procesos sobre la imagen gris; el primero será el algoritmo sobel para conseguir la extracción de los bordes de las flores y el segundo, que es un paso fundamental para el correcto

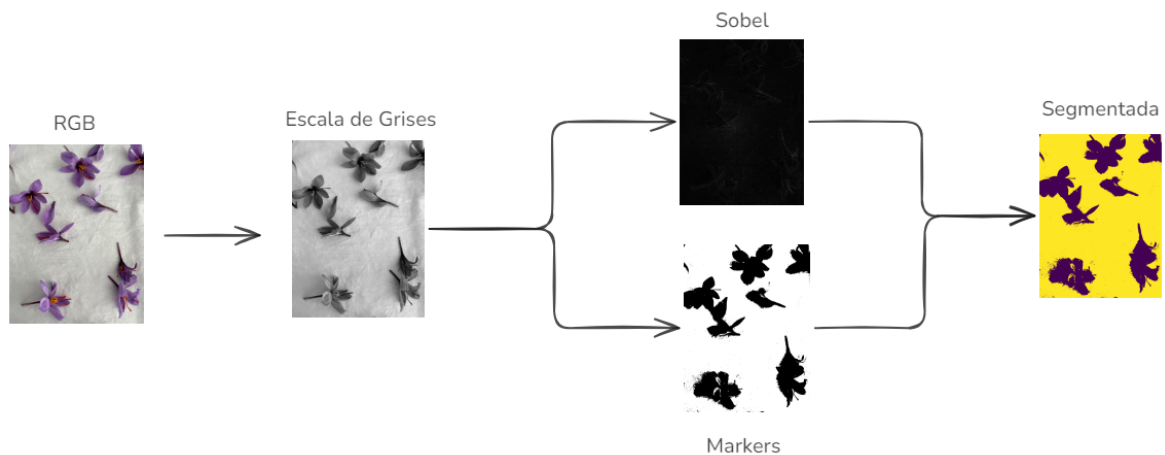


Figura 1.1: Procedimiento Segmentación Clásica

funcionamiento de watershed, consiste en la extracción de los markers. Los markers son los mínimos locales de los que parte el algoritmo. En concreto estos markers estarán representados por una imagen binaria con el mismo tamaño que la original obtenida mediante un algoritmo de thresholding [Yen et al., 1995]. A partir de ahí, hemos clasificado los píxeles que quedan por debajo del margen con un 1 y los que estén por encima con un 2. Con el resultado del sobel como mapa de elevación y con los markers hemos aplicado el algoritmo de watershed y hemos obtenido la imagen segmentada del esquema 1.1

Los resultados de aplicar este proceso sobre cuatro imágenes del dataset viene reflejado en la figura 1.2

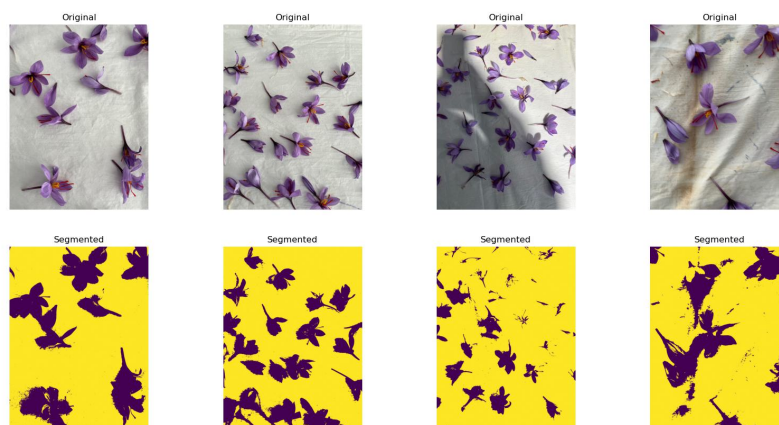


Figura 1.2: Resultados Segmentación Basada en Regiones

1.2. Segmentación con Redes Convolucionales

Como ya hemos comentado, esta segmentación está basada en el trabajo de Asako Kanezaki [Kanezaki, 2018]. En ese estudio, el objetivo es obtener un número arbitrario R de regiones a partir de una imagen no etiquetada y sin la necesidad de que se produzca ningún entrenamiento previo. Nótese que esta tarea se puede describir matemáticamente de la siguiente forma:

$$cn = f(X_n)$$

Donde X_n es el conjunto de píxeles de la imagen y f es una función $f : \mathbb{R} \rightarrow \mathbb{Z}$ que asigna a cada píxel X_i el valor de un cluster c_i .

Teniendo esto en cuenta y con el objetivo de obtener una mejor clusterización, se implementan las siguientes restricciones:

1. Los píxeles serán clasificados en función de su similaridad
2. Será preferible mantener una continuidad espacial
3. Límite al número mínimo de clusteres para evitar undersegmentation.

El primer punto se obtiene con la extracción de características que van a realizar las capas convolucionales. El algoritmo va a estar formado por M componentes, cada uno de ellos formado por una convolucion 2D, una ReLU y una batch normalization.

La segunda restricción es la preferencia de que los píxeles que se encuentren contiguos tengan más posibilidades de ser clasificados dentro del mismo cluster. Para ello, hemos utilizado el método SLIC [Achanta et al., 2012] para extraer los superpíxeles de la imagen y hemos forzado a que todos los píxeles de un superpíxel pertenezcan al mismo cluster, que será el que aparezca con mayor frecuencia.

La tercera restricción es simplemente un límite inferior al número único de clústeres de la segmentación, es decir, si ponemos este límite a 2, como será en nuestro caso para distinguir las hojas del fondo, en el momento en el que se predice sólo dos cluster para toda la imagen, el algoritmo convergerá y finalizará.

El esquema de la red completa, extraído del trabajo mencionado anteriormente, viene representado en la figura 1.3

Los resultados de aplicar la red sobre las mismas cuatro imágenes que las utilizadas en la visión clásica están representados en la figura 1.4

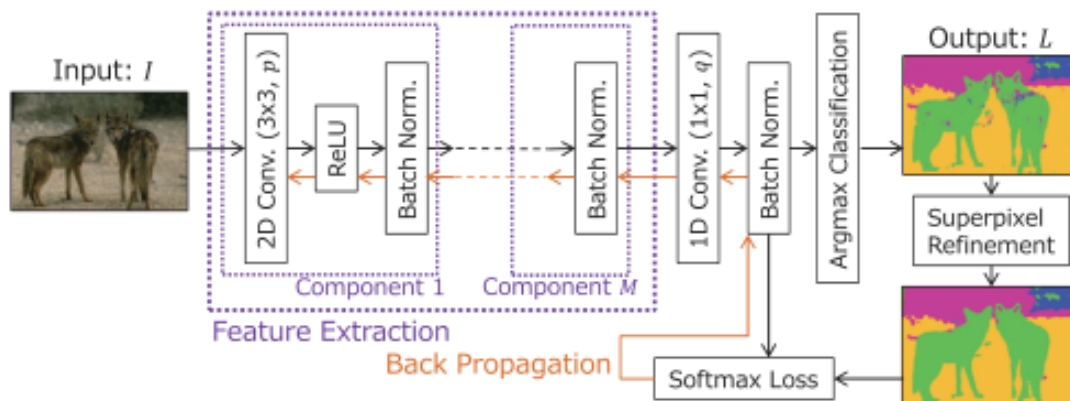


Figura 1.3: Esquema de la Red Convolutiva



Figura 1.4: Resultados Sementación Convolutiva

Como podemos observar en los resultados de la clusterización, el algoritmo ajusta mucho mejor que el anterior en la extracción de las hojas pero, al no poder definir el número final de clusters únicos, en las imágenes con peor visibilidad el algoritmo converge con un número de etiquetas mayor que 2

Esto es un problema ya que el objetivo de esta segmentación es etiquetar los datos para poder entrenar un modelo de segmentación más preciso.

Ante esta situación se proponen dos alternativas:

1. Realizar una intervención manual únicamente sobre los datos que hayan sido etiquetados con más de dos grupos. Partiríamos de la suposición de que cuando se converge en dos grupos las imágenes sí que quedan bien etiquetadas y por lo tanto que quedarían lista para el entrenamiento del modelo. El problema de esta alternativa es que si al final quedan mal etiquetadas un número elevado de imágenes el trabajo manual es muy grande.
2. Observando los resultados sobre la imagen 31 en la que, por el brillo sobre algunas hojas, se forma un grupo muy minoritario, se podría plantear hacer un postprocesamiento para que, de forma programática, se transforme ese grupo minoritario en el mayoritario o que coja un grupo por cercanía. El problema en este caso es que los outliers como la imagen 23 podrían dar un resultado muy malo.

A. Anexo 1

Referencia bibliográfica

- [Achanta et al., 2012] Achanta, R., Shaji, A., Smith, K., Lucchi, A., Fua, P., and Süsstrunk, S. (2012). Slic superpixels compared to state-of-the-art superpixel methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34.
- [Kanezaki, 2018] Kanezaki, A. (2018). Unsupervised image segmentation by backpropagation. In *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, volume 2018-April.
- [Yen et al., 1995] Yen, J. C., Chang, F. J., and Chang, S. (1995). A new criterion for automatic multilevel thresholding. *IEEE Transactions on Image Processing*, 4.