

## Miniproyecto 1 Entrega 2

Natalia Ortiz Vásquez  
202012954  
n.ortizv

María Camila Santamaría  
202015359  
m.santamariaa

Eduardo Herrera Alba  
201912865  
ej.herreraa

### 1. Cálculo de precisión y cobertura

#### 1.1. Explique con sus palabras cómo se determina un falso positivo y un falso negativo en un problema de detección y la razón detrás de este.

Un problema de detección es un problema de clasificación con máscaras binarias, es decir, este se puede evaluar realizando operaciones entre matrices de unos (1) y ceros (0), de esta manera, se logra interpretar que un falso positivo es una predicción de la clase de interés que no se acertó, mientras que un falso negativo es una anotación que no se ve reflejada en la predicción. En otras palabras, si mi clase de interés son las intensidades de valor 1, y mi anotación es 0, pero la predicción arrojada es 1, esto es un falso positivo, mientras que, si mi anotación es 1, pero la predicción es 0, esto es un caso de falso negativo.

La manera en la que se determinan los falsos positivos y falsos negativos es primeramente con el umbral de confianza, aquellos que no pasen este, serán falsos negativos. Por otro lado, si este lo sobrepasa, pasamos a calcular el índice de Jaccard y compararlo con el umbral de Jaccard, el cual me permite saber cuánto se parecen mi predicción y mi anotación. Si el índice supera al umbral, mi predicción y anotación se parecen lo suficiente, por lo que puedo decir que este es un verdadero positivo. Por último, si el índice no supera el umbral de Jaccard, se puede decir que se tiene tanto un falso positivo como un falso negativo, puesto que si bien tuve en cuenta la predicción, no es lo suficientemente parecida a la anotación, entonces la anotación no fue detectada (falso negativo) y la predicción detectó algo que no era (falso positivo).

La importancia de tener en cuenta que existen falsos positivos y falsos negativos y cómo se dan estos nos permite realizar acciones correctivas en nuestro algoritmo para así volverlo cada vez más preciso y con mayor cobertura.

#### 1.2. En este problema, ¿es preferible una alta precisión o una alta cobertura?

Creemos que es mejor tener una alta cobertura, ya que, es un problema de detección de glóbulos rojos. Se puede decir que la cobertura es la proporción de los verdaderos positivos sobre todos los objetos de mi interés. Por tanto, al tener una alta cobertura, podemos afirmar que estamos detectando la gran mayoría de nuestros objetos de interés.

### 2. Cálculo de curvas de precisión y cobertura y mAP

#### 2.1. ¿A qué parámetro es sensible la curva de precisión y cobertura y por qué?

Como pudimos observar, la curva de precisión y cobertura es sensible al umbral de Jaccard, ya que, es este umbral el que determina la cantidad de verdaderos positivos, falsos negativos y falsos positivos que existirán en nuestro problema de detección. Como se observó en el ejercicio realizado, al variar el umbral de Jaccard el área bajo la curva de precisión y cobertura cambiaba de manera significativa, lo cual, corrobora la sensibilidad de esta curva con respecto al umbral de Jaccard.

#### 2.2. ¿En cuál umbral de Jaccard se obtuvo una mayor F medida?

La F medida mayor fue de 0.97, con un **umbral de Jaccard de 0.1**, al igual que un umbral de confianza de 0.1.

#### 2.3. ¿Es preferible tener un umbral de Jaccard alto o bajo? ¿Cuáles son sus ventajas y desventajas?

Como se mencionó anteriormente, el umbral de Jaccard es aquel que me determina que tan parecidas son mi predicción y mi anotación, por tanto, si yo tengo un umbral de Jaccard alto, puedo afirmar con bastante confianza que mis verdaderos positivos son válidos. Sin embargo, al aumentar mucho este umbral, como se observó en el ejercicio realizado, la cantidad de verdaderos positivos disminuye

demasiado, lo cual ya convertiría en el algoritmo en poco eficiente y obsoleto.

#### **2.4. ¿Qué implica tener la máxima F medida en un bajo umbral de confianza?**

Esto implicaría un aumento de los Falsos positivos, ya que, se lograría tener una cobertura alta, pero una muy baja precisión. Esto se podría traducir en un algoritmo poco confiable y poco eficaz.

#### **2.5. Determine una metodología para obtener el mejor umbral de Jaccard sin comprometer la generalización del modelo**

La mejor manera de determinar el umbral de Jaccard es evaluando diferentes combinaciones para el umbral de Jaccard y el umbral de confianza, tal y como se realizó en el código. Variar el umbral de Jaccard y el umbral de confianza y para cada combinación de estos calcular las métricas de precisión y cobertura, o en este caso la F-medida. Al terminar de evaluar diferentes umbrales, hacer una comparación de cuál maximizó la F-medida y así determinar qué umbral logra maximizar la precisión o la cobertura según mi modelo lo requiera.

### **3. Detección mediante segmentación**

#### **3.1. ¿Cuáles son los problemas más importantes con cada metodología?**

Para la metodología de la grilla vemos que el problema más grande es la falta de precisión, pues esta divide la imagen en cuadrantes fijos y clasifica cada cuadrante en función de la cantidad de píxeles "True". Esto puede llevar a falsos positivos o falsos negativos, especialmente si la forma o el tamaño de los glóbulos blancos varían en la imagen. Además, el umbral de área utilizado puede ser difícil de ajustar adecuadamente pues si se elige un umbral demasiado bajo, puede haber falsos positivos, y si se elige un umbral demasiado alto, puede haber falsos negativos. Por último, dado que la grilla no puede adaptarse a las variaciones de los glóbulos blancos, el modelo podría perder detecciones si los glóbulos blancos no se alinean perfectamente con la grilla.

Respecto al método de ventana deslizante, este es más preciso pero más "costoso" y lento porque el modelo tiene muchas más ventanas que analizar. Además, se necesita ajustar cuidadosamente el tamaño y paso de la ventana deslizante para garantizar una detección adecuada. Si la ventana es demasiado pequeña o el paso demasiado grande, es posible que se pierdan glóbulos blancos pequeños. Por último, si los glóbulos se encuentran parcialmente fuera de la ventana deslizante, la metodología podría no detectarlos correctamente. Esto resulta problemático si los glóbulos son demasiado grandes o si estos se encuentran cerca al borde

de la imagen. En conjunto, las dos metodologías presentan la problemática de que solo detectan un glóbulo blanco a la vez.

#### **3.2. ¿A qué parámetro es sensible ambas metodologías?**

Ambas metodologías son sensibles al parámetro del umbral de confianza que se utiliza para determinar si una región o cuadrante se clasifica como positiva o negativa dependiendo de la cantidad de píxeles "True" dentro de esa región. El umbral escogido controla la sensibilidad de la detección, ya que determina cuántos píxeles como mínimo deben considerarse como parte de un glóbulo para que este se detecte. Si el umbral es demasiado alto, se pueden pasar por alto glóbulos blancos y si el umbral es muy bajo, se pueden generar falsos positivos. Es por esto que ajustar el umbral de manera adecuada es crucial para obtener los mejores resultados.

#### **3.3. ¿Qué debemos cambiar si deseamos detectar más de un glóbulo blanco en una única imagen?**

En primer lugar sería importante cambiar el umbral para que este sea más sensible y así se detecten regiones más pequeñas de glóbulos blancos. Luego, sería importante diseñar una parte del código que tenga la habilidad de fusionar detecciones múltiples para evitar duplicados. Se podrían utilizar diferentes técnicas para unir o eliminar detecciones que sean redundantes. Además, sería importante considerar de qué manera se manejarían las detecciones que se superponen para que el modelo cuente los glóbulos blancos por separado. Por último, es importante utilizar métricas como la F-medida para evaluar la calidad de las detecciones de múltiples glóbulos blancos.

#### **3.4. Proponga posibles mejoras para cada metodología.**

Una posible mejora para la metodología de la grilla sería cambiar la grilla con cuadrantes de tamaño fijo por una grilla que tenga la capacidad de adaptarse a la variabilidad en el tamaño de los glóbulos blancos.

Por otro lado, en la metodología de la ventana deslizante, se podría implementar un enfoque en el que, una vez que el modelo encuentre un máximo, realice una búsqueda en sus cercanías con un paso más pequeño que el utilizado inicialmente. Esto permitiría una búsqueda más detallada y precisa para confirmar si el máximo encontrado es realmente el máximo de la imagen que se está analizando.

**Realizado en L<sup>A</sup>T<sub>E</sub>X.**