

Práctica 2 VA

Pablo García López - pablo.glopez@udc.es

1. Esquema Global

1. Preprocesado de la imagen de hematoxilina.
2. Segmentación inicial de regiones nucleares.
3. Separación de núcleos individuales (segmentación de instancias).
4. Conteo y medición del área de cada núcleo.
5. Evaluación cuantitativa de los resultados con las siguientes métricas:
 - Coeficiente Dice, para medir el solapamiento instancia a instancia entre la segmentación predicha y el ground truth.
 - Error relativo medio en el conteo de núcleos.
 - Error relativo medio en el área media de los núcleos.

2. Metodologías Consideradas

Durante el análisis preliminar del conjunto de datos se observó una marcada heterogeneidad morfológica de tamaño, forma y grado de contacto entre los núcleos celulares. En particular, se identificaron dos regímenes morfológicos predominantes.

El primer régimen, denominado *Set 1*, está formado por imágenes con núcleos pequeños, compactos y relativamente homogéneos, con un buen contraste respecto al fondo. Este subconjunto está compuesto por 9 imágenes y representa un escenario favorable para métodos basados en umbralización local y operaciones morfológicas simples.

El segundo régimen, denominado *Set 2*, está formado por 10 imágenes con núcleos de mayor tamaño, frecuentemente irregulares y con un alto grado de contacto entre ellos, lo que dificulta su separación mediante técnicas morfológicas básicas. Las 11 imágenes restantes presentan características mixtas y se emplearon únicamente para la evaluación global de los métodos propuestos.

3. Metodología para el Set 1

3.1. Iteración 1: Umbralización local

Se aplica un método de segmentación simple basado en umbralización local. La imagen de hematoxilina se convierte a escala de grises y se suaviza con un filtro gaussiano ($\sigma = 3,5$).

3.1.1. Motivación

- Proporcionar una primera aproximación sencilla a la segmentación de regiones nucleares a partir del canal de hematoxilina.
- Analizar el comportamiento de un método de umbralización local en un escenario con variabilidad de intensidad y densidad celular.

3.1.2. Evaluación de resultados

Cuadro 1: Resultados Iteración 1.a: Umbralización local (Set 1)

Métrica	Valor medio
Coeficiente Dice	$0,5372 \pm 0,0470$
Error relativo medio en el conteo (%)	18,24
Error relativo medio en el área media (%)	39,58

Cuadro 2: Resultados Iteración 1.a: Umbralización local (evaluación global)

Métrica	Valor medio
Coeficiente Dice	$0,4017 \pm 0,1278$
Error relativo medio en el conteo (%)	25,71
Error relativo medio en el área media (%)	62,95

3.1.3. Conclusiones

La umbralización local permite identificar de forma fiable el interior de los núcleos, pero las regiones segmentadas presentan un tamaño inferior al real, lo que se traduce en una subestimación sistemática del área media y en un coeficiente Dice reducido. Esta observación, corroborada mediante inspección visual de las máscaras obtenidas, pone de manifiesto la necesidad de introducir mecanismos que permitan recuperar los bordes nucleares perdidos durante la umbralización.

3.2. Iteración 2: Umbralización local + dilatación

Se parte nuevamente de la imagen de hematoxilina en escala de grises, que se suaviza mediante un filtro gaussiano ($\sigma = 3,5$). A partir de la máscara binaria obtenida mediante umbralización local, se aplica una operación morfológica de dilatación con el objetivo de corregir la subestimación del área causada por la umbralización.

3.2.1. Motivación

- Compensar la infraestimación del tamaño nuclear observada en la iteración anterior.
- Evaluar si la recuperación de los contornos nucleares mediante una dilatación morfológica mejora la correspondencia entre las regiones segmentadas y los núcleos reales.

3.2.2. Evaluación de resultados

Cuadro 3: Resultados Iteración 2: Umbralización local con dilatación (Set 1)

Métrica	Valor medio
Coeficiente Dice	$0,6188 \pm 0,0515$
Error relativo medio en el conteo (%)	19,97
Error relativo medio en el área media (%)	18,33

Cuadro 4: Resultados Iteración 2: Umbralización local con dilatación (evaluación global)

Métrica	Valor medio
Coeficiente Dice	$0,4837 \pm 0,1279$
Error relativo medio en el conteo (%)	23,74
Error relativo medio en el área media (%)	44,84

3.2.3. Conclusiones

La dilatación morfológica permite recuperar parte de los bordes nucleares perdidos en la iteración anterior, incrementando el tamaño de las regiones segmentadas y mejorando su correspondencia con los núcleos reales. No obstante, al expandir las regiones segmentadas, la dilatación incrementa la probabilidad de fusión entre núcleos adyacentes, lo que limita su generalización fuera del *Set 1*.

4. Metodología para el Set 2

4.1. Iteración 3: Segmentación adaptada y separación de instancias

Las imágenes del *Set 2* presentan núcleos de mayor tamaño, con formas irregulares y un alto grado de contacto entre ellos, lo que reduce la eficacia de las estrategias empleadas en el *Set 1*. Para este régimen se ajustan los parámetros de suavizado y umbralización local, seleccionando escalas espaciales mayores acordes al tamaño medio de los núcleos del *Set 2*.

4.1.1. Motivación

- Adaptar la segmentación a la escala espacial de los núcleos del *Set 2*.

4.1.2. Evaluación de resultados

Cuadro 5: Resultados Iteración 3 (Set 2)

Métrica	Valor medio
Coeficiente Dice	$0,5135 \pm 0,0711$
Error relativo medio en el conteo (%)	11,90
Error relativo medio en el área media (%)	23,73

Cuadro 6: Resultados Iteración 3 (evaluación global)

Métrica	Valor medio
Coeficiente Dice	$0,511 \pm 0,077$
Error relativo medio en el conteo (%)	22,60
Error relativo medio en el área media (%)	65,79

4.1.3. Conclusiones

La Iteración 3 mejora de forma notable el conteo y la estimación del área media en el *Set 2*, confirmando la necesidad de adaptar la escala de análisis y de introducir mecanismos explícitos de separación de instancias. Sin embargo, estos parámetros no generalizan adecuadamente al conjunto completo de imágenes, lo que pone de manifiesto las limitaciones de los métodos basados en reglas fijas ante una elevada heterogeneidad morfológica.

4.2. Iteración 4: Segmentación + watershed

A partir de la segmentación obtenida en la Iteración 3, se introduce un método explícito de separación de instancias basado en watershed controlado por marcadores internos. La máscara binaria obtenida tras la umbralización adaptada y la dilatación se utiliza como región de interés sobre la que se calcula la transformada de distancia euclídea.

Los máximos locales de la transformada de distancia se emplean como marcadores internos, permitiendo inicializar el algoritmo watershed de forma robusta. De este modo, la separación de núcleos se realiza únicamente en regiones donde existe contacto real entre instancias.

4.2.1. Motivación

- Separar núcleos en contacto sin alterar la segmentación base de área.
- Mejorar el conteo de núcleos en escenarios de alta densidad celular.
- Evaluar el impacto de un watershed controlado frente a la segmentación directa.

4.2.2. Evaluación de resultados

Cuadro 7: Resultados Iteración 4 (Set 2)

Métrica	Valor medio
Coeficiente Dice	$0,5535 \pm 0,0772$
Error relativo medio en el conteo (%)	8,30
Error relativo medio en el área media (%)	21,65

Cuadro 8: Resultados Iteración 4 (Set completo)

Métrica	Valor medio
Coeficiente Dice	$0,5634 \pm 0,0866$
Error relativo medio en el conteo (%)	19,70
Error relativo medio en el área media (%)	54,59

4.2.3. Conclusiones

La introducción del watershed controlado por marcadores mejora de forma significativa el conteo de núcleos y la estimación del área media en el *Set 2*.

Considerando exclusivamente la evaluación global sobre el conjunto completo de imágenes, la Iteración 4 presenta el mejor compromiso entre solapamiento, conteo y estimación de área, constituyendo la solución más robusta cuando se aplica de forma uniforme a todas las imágenes.

5. Conclusiones finales

Los resultados obtenidos muestran que la segmentación de núcleos celulares en imágenes histológicas presenta problemáticas diferenciadas en función de su morfología. Mientras que métodos basados en umbralización local y operaciones morfológicas simples resultan adecuados para imágenes con núcleos pequeños y homogéneos, la presencia de núcleos grandes y en contacto requiere estrategias específicas de separación de instancias.

Estas observaciones evidencian las limitaciones de los métodos clásicos de Visión Artificial frente a conjuntos de datos altamente heterogéneos y sugieren que enfoques basados en aprendizaje automático podrían ofrecer una mayor capacidad de generalización.