



Segmentación semántica en imágenes de microscopía utilizando redes neuronales

Autor:

Lic. Pedro Miguel Pérez

Director:

Nombre del Director (pertenencia)

*Esta planificación fue realizada en el curso de Gestión de proyectos
entre el 24 de junio de 2021 y el 18 de agosto de 2021.*

Índice

1. Descripción técnica-conceptual del proyecto a realizar	5
2. Identificación y análisis de los interesados	6
3. Propósito del proyecto	7
4. Alcance del proyecto	7
5. Supuestos del proyecto.	7
6. Requerimientos	8
7. Historias de usuarios (<i>Product backlog</i>).	8
8. Entregables principales del proyecto	8
9. Desglose del trabajo en tareas	9
10. Diagrama de Activity On Node.	9
11. Diagrama de Gantt	10
12. Presupuesto detallado del proyecto	13
13. Gestión de riesgos	13
14. Gestión de la calidad	14
15. Procesos de cierre	15

Registros de cambios

Revisión	Detalles de los cambios realizados	Fecha
0	Creación del documento	24 de junio de 2021
1	Se completa hasta el punto 5 inclusive	08 de junio de 2021

Acta de constitución del proyecto

Buenos Aires, 24 de junio de 2021

Por medio de la presente se acuerda con el Lic. Pedro Miguel Pérez que su Trabajo Final de la Carrera de Especialización en Inteligencia Artificial se titulará “Segmentación semántica en imágenes de microscopía utilizando redes neuronales” que consistirá esencialmente en la implementación de una red neuronal para el procesamiento de imágenes de microscopía, y tendrá un presupuesto preliminar estimado de 600 hs de trabajo, con fecha de inicio 24 de junio de 2021 y fecha de presentación pública 15 de mayo de 2022.

Se adjunta a esta acta la planificación inicial.

Ariel Lutenberg
Director posgrado FIUBA

Mara Alderete
Instituto de Nanosistemas UNSAM

Nombre del Director
Director del Trabajo Final

1. Descripción técnica-conceptual del proyecto a realizar

Los esferoides son un modelo biológico de cultivo celular con estructura tridimensional. Los mismos son utilizados ampliamente en biología celular por poseer características similares al tejido del que provienen. Este tipo de cultivo constituye una herramienta atractiva dado que permite concluir que las condiciones de mantenimiento de las células en el cultivo son similares a las que se observan en un tejido de un organismo vivo.

Es posible generar esferoides mediante distintas metodologías, siendo una de las más utilizadas la denominada “gota pendiente”, que se basa en crear gotas de una suspensión celular sobre la cara interior de una superficie de vidrio o plástico. En cada gota creada las células forman un agregado con una estructura tridimensional similar a una esfera.

Las aplicaciones para las que han sido utilizados los esferoides son diversas, entre las que se destacan el screening de drogas con el objetivo de determinar aquellas más eficientes para combatir el crecimiento tumoral; la respuesta del esferoide al tratamiento se analiza midiendo dos parámetros predictivos claves: (I) variación de volumen y (II) cambios en la forma. Para ello, se suele registrar ambos indicadores a lo largo del tiempo realizando mediciones sobre imágenes obtenidas mediante microscopía de campo claro, lo cual conlleva un tiempo y esfuerzo considerable para el operador en el caso que realice el análisis de forma manual.

Actualmente existe la necesidad de poder automatizar la medición de diferentes características en imágenes microscópicas de esferoides. El análisis manual realizado es propenso a errores, y a su vez añade el sesgo del operador que realiza la medición. Debido a esto, un método de procesamiento automático mejoraría la calidad y trazabilidad de los resultados y disminuiría considerablemente el tiempo de procesamiento. En la Figura 1 se muestra la secuencia de análisis y cuantificación de medidas en un esferoide.

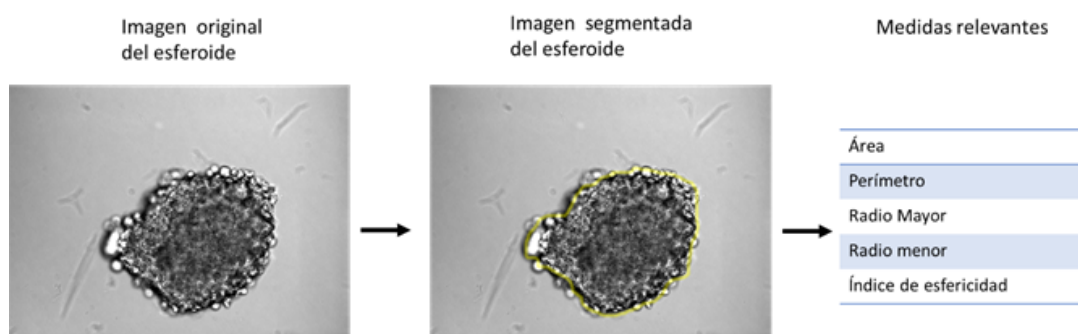


Figura 1. Secuencia de análisis de imágenes y cuantificación en un esferoide.

Las disciplinas computacionales centradas en inteligencia artificial pueden aportar estrategias para facilitar el análisis y proponer soluciones a una variedad de problemas biológicos. Específicamente en este caso, contar con un algoritmo que pueda realizar este análisis automáticamente a partir de la segmentación de los esferoides presentes en las imágenes sería sumamente interesante para diversos grupos de investigación en este campo. Si bien se han realizado esfuerzos en la comunidad científica para avanzar con desarrollos en este sentido, los softwares generados no reconocen adecuadamente todos los formatos de imagen almacenados por distintos microscopios. Adicionalmente, la presencia de más de un esferoide por campo, o de bordes difusos en la imagen, generan dificultades para el adecuado funcionamiento de la rutina, introduciendo errores en la segmentación.

El objetivo de este trabajo es desarrollar un software capaz de segmentar y caracterizar automáticamente esferoides en cultivo a partir de imágenes obtenidas con microscopía de transmisión en campo claro, y así extraer medidas relevantes como área, radio e índice de esfericidad.

Las redes neuronales, en especial las redes neuronales convolucionales (Convolution Neural Network en inglés - CNN), han demostrado su amplio potencial en el procesamiento de imágenes para resolver problemas de clasificación y segmentación de objetos. El objetivo principal de este trabajo es poder aplicar técnicas de aprendizaje automático, en particular redes neuronales convolucionales, para procesar imágenes automáticamente y así poder segmentar y cuantificar esferoides en las mismas. En este trabajo el principal desafío es crear un sistema automático para el procesamiento de imágenes en un entorno real contando con un número limitado de imágenes para el entrenamiento.

El esquema de trabajo propuesto consiste en tres fases: generación del dataset, entrenamiento del modelo y aplicación del modelo. En la etapa de generación del dataset, a partir de un conjunto de imágenes reales se realizará un proceso de segmentación manual en la cual se extraerá el área de interés o ROI (region of interest en inglés), las imágenes segmentadas se separarán en dos sets, uno de ellos será utilizado para el entrenamiento y el otro para la evaluación del modelo. En la aplicación del modelo, se plantea una aplicación web que permitirá ingresar una imagen al modelo para lograr la predicción del área de interés o ROI del esferoide, a partir de esto se realizará una cuantificación de las medidas de interés, perímetro, radios e índice de esfericidad. En la Figura 2 se presenta el esquema de trabajo propuesto.

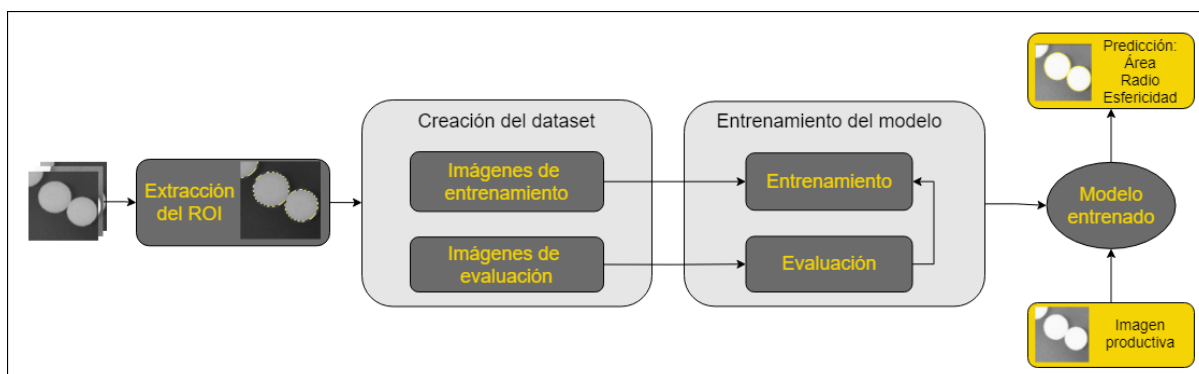


Figura 2. Esquema de trabajo propuesto.

2. Identificación y análisis de los interesados

Rol	Nombre y Apellido	Organización	Puesto
Cliente	Lic. Mara Alderete	Instituto de Nanosistemas UNSAM	Coordinadora UNSAM
Responsable	Lic. Pedro Miguel Pérez	FIUBA	Alumno
Colaboradores	Lic. Marina Simian	CONICET	Investigadora
Orientador	Nombre del Director	pertenencia	Director Trabajo final
Usuario final	Operador	UNSAM	Analista

- Colaboradores: Marina Simian será de gran ayuda para la creación de los dataset necesarios para entrenar y validar el modelo.

- Usuario final: Los analistas del laboratorio serán de gran ayuda para validar el modelo en un entorno productivo.

3. Propósito del proyecto

El propósito de este proyecto es diseñar e implementar una solución basada en una arquitectura de red neuronal convolucional que permita determinar características morfológicas aplicando técnicas de detección y segmentación eficiente en imágenes de microscopía.

4. Alcance del proyecto

El desarrollo del proyecto incluye:

- Desarrollar una herramienta que permita determinar el tamaño de esferoides en una imagen de microscopía utilizando redes neuronales convolucionales (CNN).
- Estudiar y comparar diferentes arquitecturas de redes neuronales para evaluar su rendimiento. Se plantea explorar las siguientes arquitecturas (U-NET, FastFCN y Mask R-CNN).
- Crear un dataset de imágenes que permita el entrenamiento de la red.
- Transformar y adaptar el dataset utilizado con el fin de incrementar el número de muestras y así poder aumentar la performance del modelo.
- Desarrollar una CNN que permita determinar la distribución del tamaño de partículas en dos clases de imágenes: esferoides aislados y aglomerado de los mismos.
- Evaluar la efectividad del modelo y aplicar técnicas que permitan mejorar el rendimiento de la red aplicando técnicas de optimización de los hiper parámetros.
- Escalar el modelo para el procesamiento de imágenes de naturaleza distinta a los esferoides, como por ejemplo imágenes de nanopartículas.
- Desarrollar un servicio web básico, donde se podrá hacer uso de la red neuronal para analizar las imágenes de esferoides.

5. Supuestos del proyecto

Para el desarrollo del presente proyecto se supone que:

- Se dispondrá un conjunto de imágenes previamente segmentadas para realizar el entrenamiento del modelo.
- El cliente deberá colaborar en la recolección y segmentación manual de las imágenes que serán utilizadas para el entrenamiento del modelo.
- El entrenamiento del modelo será realizado sobre la plataforma Colab.

6. Requerimientos

1. Requerimientos funcionales

- 1.1. El sistema debe permitir seleccionar una imagen desde una interfaz web para ser procesada.
- 1.2. El algoritmo debe detectar un esferoide dentro de la imagen procesada y mostrar la segmentación realizada.
- 1.3. El algoritmo debe medir el esferoide detectado y brindar información del área, radio y perímetro.
- 1.4. El sistema debe ser fácil de utilizar.

2. Requerimientos no funcionales.

- 2.1. El sistema debe permitir procesar imágenes en diferentes formatos.
- 2.2. El sistema debe poseer un rango de error menor al 5 por ciento con respecto al dataset de evaluación
- 2.3. El sistema debe poseer una interfaz web para procesar las imágenes.
- 2.4. El sistema debe poder procesar una imagen en un rango de tiempo aceptable.

7. Historias de usuarios (*Product backlog*)

Descripción: En esta sección se deben incluir las historias de usuarios y su ponderación (*history points*). Recordar que las historias de usuarios son descripciones cortas y simples de una característica contada desde la perspectiva de la persona que desea la nueva capacidad, generalmente un usuario o cliente del sistema. La ponderación es un número entero que representa el tamaño de la historia comparada con otras historias de similar tipo.

El formato propuesto es: como [rol] quiero [tal cosa] para [tal otra cosa].”

Se debe indicar explícitamente el criterio para calcular los *story points* de cada historia

8. Entregables principales del proyecto

Los entregables del proyecto son (ejemplo):

- Manual de uso
- Diagrama de circuitos esquemáticos
- Código fuente del firmware
- Diagrama de instalación
- Informe final
- etc...

9. Desglose del trabajo en tareas

El WBS debe tener relación directa o indirecta con los requerimientos. Son todas las actividades que se harán en el proyecto para dar cumplimiento a los requerimientos. Se recomienda mostrar el WBS mediante una lista indexada:

1. Grupo de tareas 1

- 1.1. Tarea 1 (tantas hs)
- 1.2. Tarea 2 (tantas hs)
- 1.3. Tarea 3 (tantas hs)

2. Grupo de tareas 2

- 2.1. Tarea 1 (tantas hs)
- 2.2. Tarea 2 (tantas hs)
- 2.3. Tarea 3 (tantas hs)

3. Grupo de tareas 3

- 3.1. Tarea 1 (tantas hs)
- 3.2. Tarea 2 (tantas hs)
- 3.3. Tarea 3 (tantas hs)
- 3.4. Tarea 4 (tantas hs)
- 3.5. Tarea 5 (tantas hs)

Cantidad total de horas: (tantas hs)

Se recomienda que no haya ninguna tarea que lleve más de 40 hs.

10. Diagrama de Activity On Node

Armar el AoN a partir del WBS definido en la etapa anterior.

Indicar claramente en qué unidades están expresados los tiempos. De ser necesario indicar los caminos semicríticos y analizar sus tiempos mediante un cuadro. Es recomendable usar colores y un cuadro indicativo describiendo qué representa cada color, como se muestra en el siguiente ejemplo:



Figura 3. Diagrama en *Activity on Node*

11. Diagrama de Gantt

Existen muchos programas y recursos *online* para hacer diagramas de gantt, entre los cuales destacamos:

- Planner
- GanttProject
- Trello + *plugins*. En el siguiente link hay un tutorial oficial:
<https://blog.trello.com/es/diagrama-de-gantt-de-un-proyecto>
- Creately, herramienta online colaborativa.
<https://creately.com/diagram/example/ieb3p3ml/LaTeX>
- Se puede hacer en latex con el paquete *pgfgantt*
<http://ctan.dcc.uchile.cl/graphics/pgf/contrib/pgfgantt/pgfgantt.pdf>

Pegar acá una captura de pantalla del diagrama de Gantt, cuidando que la letra sea suficientemente grande como para ser legible. Si el diagrama queda demasiado ancho, se puede pegar primero la “tabla” del Gantt y luego pegar la parte del diagrama de barras del diagrama de Gantt.

Configurar el software para que en la parte de la tabla muestre los códigos del EDT (WBS).
Configurar el software para que al lado de cada barra muestre el nombre de cada tarea.
Revisar que la fecha de finalización coincida con lo indicado en el Acta Constitutiva.

En la figura 4, se muestra un ejemplo de diagrama de gantt realizado con el paquete de *pgfgantt*. En la plantilla pueden ver el código que lo genera y usarlo de base para construir el propio.



Figura 4. Diagrama de gantt de ejemplo



Figura 5. Ejemplo de diagrama de Gantt rotado

12. Presupuesto detallado del proyecto

Si el proyecto es complejo entonces separarlo en partes:

- Un total global, indicando el subtotal acumulado por cada una de las áreas.
- El desglose detallado del subtotal de cada una de las áreas.

IMPORTANTE: No olvidarse de considerar los **COSTOS INDIRECTOS**.

COSTOS DIRECTOS			
Descripción	Cantidad	Valor unitario	Valor total
SUBTOTAL			
COSTOS INDIRECTOS			
Descripción	Cantidad	Valor unitario	Valor total
SUBTOTAL			
TOTAL			

13. Gestión de riesgos

a) Identificación de los riesgos (al menos cinco) y estimación de sus consecuencias:

Riesgo 1: detallar el riesgo (riesgo es algo que si ocurre altera los planes previstos de forma negativa)

- Severidad (S): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S).
- Probabilidad de ocurrencia (O): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2:

- Severidad (S):
- Ocurrencia (O):

Riesgo 3:

- Severidad (S):

■ Ocurrecia (O):

b) Tabla de gestión de riesgos: (El RPN se calcula como $RPN=S \times O$)

Riesgo	S	O	RPN	S*	O*	RPN*

Criterio adoptado: Se tomarán medidas de mitigación en los riesgos cuyos números de RPN sean mayores a...

Nota: los valores marcados con (*) en la tabla corresponden luego de haber aplicado la mitigación.

c) Plan de mitigación de los riesgos que originalmente excedían el RPN máximo establecido:

Riesgo 1: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación). Nueva asignación de S y O, con su respectiva justificación: - Severidad (S): mientras más severo, más alto es el número (usar números del 1 al 10). Justificar el motivo por el cual se asigna determinado número de severidad (S). - Probabilidad de ocurrencia (O): mientras más probable, más alto es el número (usar del 1 al 10). Justificar el motivo por el cual se asigna determinado número de (O).

Riesgo 2: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

Riesgo 3: plan de mitigación (si por el RPN fuera necesario elaborar un plan de mitigación).

14. Gestión de la calidad

Para cada uno de los requerimientos del proyecto indique:

- Req #1: copiar acá el requerimiento.
 - Verificación para confirmar si se cumplió con lo requerido antes de mostrar el sistema al cliente. Detallar
 - Validación con el cliente para confirmar que está de acuerdo en que se cumplió con lo requerido. Detallar

Tener en cuenta que en este contexto se pueden mencionar simulaciones, cálculos, revisión de hojas de datos, consulta con expertos, mediciones, etc. Las acciones de verificación suelen considerar al entregable como “caja blanca”, es decir se conoce en profundidad su funcionamiento interno. En cambio, las acciones de validación suelen considerar al entregable como “caja negra”, es decir, que no se conocen los detalles de su funcionamiento interno.

15. Procesos de cierre

Establecer las pautas de trabajo para realizar una reunión final de evaluación del proyecto, tal que contemple las siguientes actividades:

- Pautas de trabajo que se seguirán para analizar si se respetó el Plan de Proyecto original:
- Indicar quién se ocupará de hacer esto y cuál será el procedimiento a aplicar.
- Identificación de las técnicas y procedimientos útiles e inútiles que se emplearon, y los problemas que surgieron y cómo se solucionaron: - Indicar quién se ocupará de hacer esto y cuál será el procedimiento para dejar registro.
- Indicar quién organizará el acto de agradecimiento a todos los interesados, y en especial al equipo de trabajo y colaboradores: - Indicar esto y quién financiará los gastos correspondientes.