



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
**POSGRADO EN CIENCIA E INGENIERÍA DE LA COMPUTACIÓN**

**IDENTIFICACIÓN AUTOMÁTICA DE ACTIVIDAD NEURONAL EN**  
**IMAGENOLÓGÍA DE CALCIO**

**T E S I S**

Que para optar por el grado de:

**Maestro en Ciencia e Ingeniería de la Computación**

Presenta:

**Fernando González Colín**

Tutor principal:

Dr. Boris Escalante Ramírez

Laboratorio Avanzado de Procesamiento de Imágenes - UNAM

Ciudad Universitaria, Cd.Mx., Mayo 2023



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

*A mis padres y a Amairani mi más sincera gratitud.*

## Agradecimientos

---

Agradezco a la Universidad Nacional Autónoma de México y al Posgrado en Ciencia e Ingeniería de la Computación por la formación recibida, la cual va más allá de la técnica y me ha ofrecido la oportunidad de contribuir al desarrollo de mi familia, mi sociedad y mi nación.

Agradezco a mis profesores por el conocimiento, la inspiración y el deseo de saber más que me han transmitido con diligencia.

Agradezco al Dr. Boris Escalante Ramirez y a la Dra. Jimena Olveres Montiel por el acompañamiento, el apoyo, la confianza, la paciencia y el empuje para realizar el trabajo hecho desde hace ya cinco años, como resultado se han producido dos tesis y un artículo de conferencia.

Agradezco al Dr. José Bargas, al Dr. Miguel Serrano Reyes y al Instituto de Fisiología Celular por el apoyo y la colaboración, trabajar con ustedes me ayudó a ampliar mi visión más allá de la ingeniería y de la computación.

Agradezco a mis padres por su apoyo y cariño incondicional, por creer en mí aún cuando ni yo me siento capaz de lograr algo. Gracias por la maravillosa educación que recibí en casa, tuve la suerte de crecer entre ustedes, por lo que mis logros son sus logros.

Agradezco a Amairani por su cariño, calidez, apoyo y comprensión, por acompañarme en este proceso y por la inspiración que me brinda aún sin saberlo.

Este trabajo de tesis recibió apoyo de los proyectos UNAM PAPIIT IV100420 y TA101121 y de una beca de nivel maestría del Consejo Nacional de Ciencia y Tecnología.

## Resumen

---

Comprender el funcionamiento de la mente humana ha sido fuente de inspiración de muchas investigaciones y esfuerzos científicos. Como es natural, la búsqueda de su comprensión requiere de la interpretación de observaciones, en particular del tejido nervioso; el cuál está formado por miles de millones de neuronas, células especializadas en la generación y propagación de señales eléctricas, de cuyas interacciones se especula que emergen tanto funciones sensoriales como cognitivas. Así, su comprensión podría conducir a tratamientos para desórdenes neuronales como la enfermedad de Alzheimer o epilepsia.

Es necesario analizar las observaciones para comprender los mecanismos detrás del funcionamiento de las neuronas. Las técnicas basadas en imagenología de calcio han permitido realizar observaciones de cientos de neuronas simultáneamente y con resolución de célula única. Sin embargo, este tipo de técnicas genera grandes volúmenes de datos en forma de vídeos que deben ser examinados meticulosamente por los investigadores para, en primer lugar, identificar neuronas funcionales. Esta tarea suele ser extenuante y propensa a errores humanos, por lo que existe la necesidad de métodos automatizados para su realización.

En la literatura se encuentran propuestas que intentan resolver esta tarea, sin embargo, no se ha logrado con éxito y hay un consenso de que permanece como un problema abierto. La dificultad del reto está relacionado a las características de los vídeos y a la naturaleza de la tarea, ya que se requiere identificar patrones tanto en el espacio como el tiempo; con técnicas de procesamiento de imágenes digitales se han hecho propuestas para el análisis espacial, sin embargo, reconocer patrones en el tiempo en este tipo de datos es complicado debido al volumen de los mismos. Computacionalmente es muy costoso procesar los vídeos como una matriz tridimensional, por lo que es

común resumir toda la observación en una sola imagen, lo cuál implica la pérdida de información temporal.

En el presente trabajo estamos proponiendo un método basado en dos tipos de redes neuronales artificiales, una tipo convolucional para análisis espacial y otra del tipo recurrente para análisis temporal. También se incluyen etapas de preprocesamiento y postprocesamiento para mejorar el rendimiento de las redes, refinar sus resultados y finalmente validarlos. Para probar nuestra propuesta se cuentan con datos recolectados del Instituto de Fisiología Celular de la UNAM (IFC - UNAM).

## **1. Introducción**

- 1.1 Definición del problema
- 1.2 Objetivos
- 1.3 Relevancia
- 1.4 Metodología
- 1.5 Estructura del documento

## **2. Conceptos generales**

- 2.1 Antecedentes de neurociencias
- 2.2 Imagenología de calcio
- 2.3 Experimento característico
- 2.4 Redes Neuronales Artificiales

## **3. Estado del arte**

## **4. Metodología**

- 4.1 Análisis del problema
- 4.2 Propuesta
- 4.3 Preprocesamiento
- 4.4 Segmentación
- 4.5 Postprocesamiento
- 4.6 Preparación de datos
- 4.7 Experimentos
- 4.8 Métricas

## **5. Resultados**

- 5.1 Discusión

## **6. Conclusiones**

- 6.1 Trabajo futuro

## **Bibliografía**

# 1. INTRODUCCIÓN

---

En la búsqueda por la comprensión del sistema nervioso resulta imprescindible la observación y análisis de extensas poblaciones de los distintos tipos de células que lo conforman, como lo son las neuronas, por lo tanto también lo son las técnicas que impulsen estos análisis (Stevenson. Ian H. et al 2011). En las últimas seis décadas el número de neuronas que es posible observar se ha duplicado cada 7 años, si continúa este crecimiento exponencial, en 200 años sería posible observar simultáneamente 100 mil millones de neuronas, prácticamente la totalidad de neuronas en un cerebro humano (Stevenson. Ian H. et al 2011). Sin embargo, observar un número mayor de estas células implica también el crecimiento de los conjuntos de datos generados para almacenar las observaciones. Por motivos que se discutirán más adelante, conforme el volumen de datos aumenta se vuelve más complicado identificar individualmente cada célula, analizar su comportamiento y las formas en que interactúan entre sí, es así como la detección automática de neuronas se vuelve relevante (Apthorpe, N. J., et al 2016). Afortunadamente, de manera simultánea las capacidades en cómputo han estado creciendo exponencialmente, haciendo posible el procesamiento de volúmenes de datos cada vez más grandes (Stevenson. Ian H. et al 2011).

## 1.1 Definición del problema

Avances en técnicas de imagenología de calcio, como microscopía de epifluorescencia, hacen posible la observación de amplias áreas de tejido cerebral vivo con buena resolución temporal y el monitoreo simultáneo de la actividad neuronal de cientos de neuronas con resolución de célula única (Badea, T., et al. 2001, Cossart, R., et al. 2005, Ferran, D., et al. 2013, Giovannucci, A., et al. 2019, Kirschbaum, E., et al. 2020, Pérez, J., et al. 2016, Serrano-Reyes, M., et al. 2020). En este tipo de observaciones, registradas en

formato de vídeo, la extracción de la ubicación individual de cada célula observada es una tarea indispensable para el análisis de dichas observaciones (Kirschbaum, E., et al. 2020). Sin embargo, esta tarea resulta difícil por complicaciones relacionadas con las características de las observaciones: baja relación señal-ruido, fondo poco uniforme, pérdida de las señales de fluorescencia debido al efecto de foto blanqueo y traslape de neuronas (Ferran, D., et al. 2013), asimismo a las condiciones de las muestras observadas: alta variabilidad en densidad, morfología y distribución, la emisión de fluorescencia no suele ser uniforme y los bordes de las células tienden a ser difusos, conduciendo a segmentaciones morfológicamente incorrectas (Baglietto, S., et al. 2018).

Los vídeos que se obtienen de este tipo de observaciones, deben ser analizados meticulosamente por neurocientíficos en búsqueda de neuronas funcionales. Una neurona funcional es aquella que presenta cambios de fluorescencia, esto se observa en los vídeos en forma de pequeñas manchas con cambios significativos de intensidad a través del tiempo. En un vídeo es posible tener cientos de pequeñas manchas parpadeantes correspondientes a cientos de células nerviosas. Esta tarea es tediosa, agotadora y propensa al error humano. Neurocientíficos experimentados necesitan revisar cada vídeo cuadro a cuadro buscando ligeros cambios de intensidad en los píxeles, esto durante una hora en promedio por cada 30 minutos de experimento. En este escenario, resulta indispensable el desarrollo de herramientas automatizadas, rápidas y confiables que apoyen a los neurocientíficos en la tarea de identificar las neuronas funcionales contenidas en los vídeos.

## **1.2 Objetivos**

Dada la complejidad de resolver el problema descrito anteriormente, se tiene por objetivo realizar un análisis de las diversas propuestas hechas en la literatura para identificar potenciales deficiencias en ellas desde una perspectiva

computacional y proponer una solución que compense tales deficiencias para identificar automáticamente neuronas funcionales en vídeos adquiridos mediante un tipo de imagenología de calcio llamada microscopía de epifluorescencia.

La solución propuesta deberá ser ajustada y evaluada con los datos proporcionados por el IFC-UNAM como parte de una colaboración multidisciplinaria para desarrollar herramientas de apoyo a las labores de los investigadores de dicho instituto.

### **1.3 Relevancia**

Responder la incógnita de cómo funciona el sistema nervioso y cómo es procesada la información en su interior es fundamental en el entendimiento de funciones cerebrales complejas, tanto cognitivas como sensoriales. Desafortunadamente, responder a esto ha sido difícil debido principalmente a la cantidad de elementos involucrados y a las dificultades de monitorear la actividad neuronal (Stosiek, C., et al. 2003). Su comprensión podría conducir al desarrollo de múltiples aplicaciones, entre las que se encuentran tratamientos para desórdenes neurológicos como la enfermedad de Alzheimer o Epilepsia (Perez-Ortega Jesus. et al. 2016) e interfaces cerebro-máquina (Stevenson, I. et al 2011).

Como se muestra en el estado del arte, la mayoría de las propuestas se han concentrado en la segmentación a nivel de imagen, de hecho la segmentación a nivel de píxel ha experimentado un enorme progreso en años recientes (Siam, M., et al. 2016), mientras que la segmentación en vídeos ha sido complicada debido en parte a la exigencia computacional que requiere procesarlos, los videos suelen ser vistos como matrices tridimensionales, dos dimensiones para el espacio y una más para el tiempo, y aplicar operaciones sobre ellas tiende ser costoso computacionalmente. Otro problema importante que ha limitado el progreso en la segmentación de vídeos es el aprovechamiento

de la información temporal, en el presente trabajo se aborda este problema con técnicas de aprendizaje profundo, las cuales han demostrado ser competitivas en problemas que requieren del análisis de secuencias.

#### **1.4 Metodología**

Al abordar el problema descrito, se analizan diferentes propuestas presentes en la literatura que han buscado resolver el mismo problema, conocer el estado del arte brinda claridad sobre las estrategias que han fallado y las que han tenido algún grado de éxito y el porqué de ello. Como se explica con más detalle en la sección 4.1, una de las limitaciones más importantes en las estrategias empleadas ha sido la pérdida de información temporal, la cuál resulta crítica en la identificación de neuronas funcionales, pues la señal de fluorescencia de una célula debe cumplir con ciertas características temporales para que la célula sea considerada como funcional.

Las redes neuronales artificiales (ANNs) han demostrado rendimientos de estado del arte en identificación de patrones en el espacio mediante elementos convolucionales, mientras que las redes recurrentes han hecho lo propio para patrones en series de tiempo. Por tal motivo es que se propone una solución basada en la combinación de estos dos tipos de ANNs. Como es común en los sistemas de reconocimiento de patrones (Jain, A. K., et al. 2000), se agrega una etapa inicial de preprocesamiento para impulsar el rendimiento de las ANNs (u otros métodos de aprendizaje). Asimismo, se agrega una etapa de postprocesamiento para refinar las segmentaciones de las ANNs y validar los resultados.

Para evaluar el desempeño de nuestra propuesta, se toman evaluaciones hechas en el estado del arte como referencia, debido a la naturaleza del problema que se intenta resolver: segmentación de dos clases en imágenes, se requieren métricas adecuadas para segmentación binaria y que compensen el desbalance de clases, como lo es el coeficiente de Sørensen-Dice. Asimismo, es

importante tener cuidado con la interpretación de métricas sensibles al desbalance, como lo son Precisión y *Recall* (se usa *Recall* en inglés debido a que las traducciones de sensibilidad o exhaustividad reflejan una idea ambigua) Por tal motivo se dedica una sección al análisis de los resultados obtenidos en dichas métricas.

## **1.5 Estructura del documento**

En la sección 2 se describen conceptos generales sobre: neurociencias, imagenología de calcio, procesamiento de imágenes digitales y redes neuronales artificiales; debido a su importancia en el contexto del presente trabajo. Asimismo, se describe un experimento característico realizado en el IFC-UNAM del tipo al que corresponden los conjuntos de datos utilizados. En la sección 3 se realiza la revisión del estado de arte, se listan algunas propuestas, comentarios relevantes hechos por los autores acerca de los retos y resultados. Posteriormente, en la sección 4 se describe a detalle la metodología del proyecto, se analiza el problema a resolver, se explica la propuesta, se describe su arquitectura, así como el proceso de adquisición y preparación de los datos utilizados. Los experimentos, métricas empleadas y una discusión sobre los resultados son presentados en la sección 5. Finalmente, en la sección 6 se presentan nuestras conclusiones y se plantea el trabajo futuro.

## 2. CONCEPTOS GENERALES

---

### 2.1 Antecedentes de neurociencias

Todos los tejidos y órganos del cuerpo humano se conforman de células y sus funciones están determinadas por las funciones especializadas de sus células y la forma en que estas interactúan (Bear, M. F., et al. 2016). Entre los diferentes órganos, el cerebro es quizás el más sofisticado y complejo del que se tenga conocimiento (Bear, M. F., et al. 2016). El cerebro de un humano adulto puede tener alrededor de 85 mil millones de un tipo de células nerviosas llamadas neuronas, estas células están altamente especializadas en la generación y propagación de señales eléctricas (Dayan, P., et al. 2005) y se encuentran interconectadas en sistemas llamados circuitos neuronales, de los que depende nuestra percepción del mundo, nuestra atención y el control de nuestras acciones (Kandel, E. R., et al. 2013). La comprensión del funcionamiento del cerebro requiere del entendimiento de cómo las neuronas funcionan individualmente (Bear, M. F., et al. 2016) y de cómo se organizan y comunican entre ellas (Kandel, E. R., et al. 2013).

Una neurona se compone de cuatro partes morfológicas bien definidas: i) cuerpo celular o soma, ii) axón, iii) terminales sinápticas y iv) dendritas. Cada una de estas partes desempeñan un rol en la generación y propagación de señales eléctricas (Kandel, E. R., et al. 2013).

**El soma** es la parte semiesférica y central de la célula y mide aproximadamente 20  $\mu\text{m}$  de diámetro. En su interior se encuentran los mismos organelos que se encuentran en todas las células animales (Bear, M. F., et al. 2016) y se mantienen aislados del exterior por la membrana de la neurona. También, es el centro metabólico de la neurona y da origen a las dendritas y al axón (Kandel, E. R., et al. 2013).

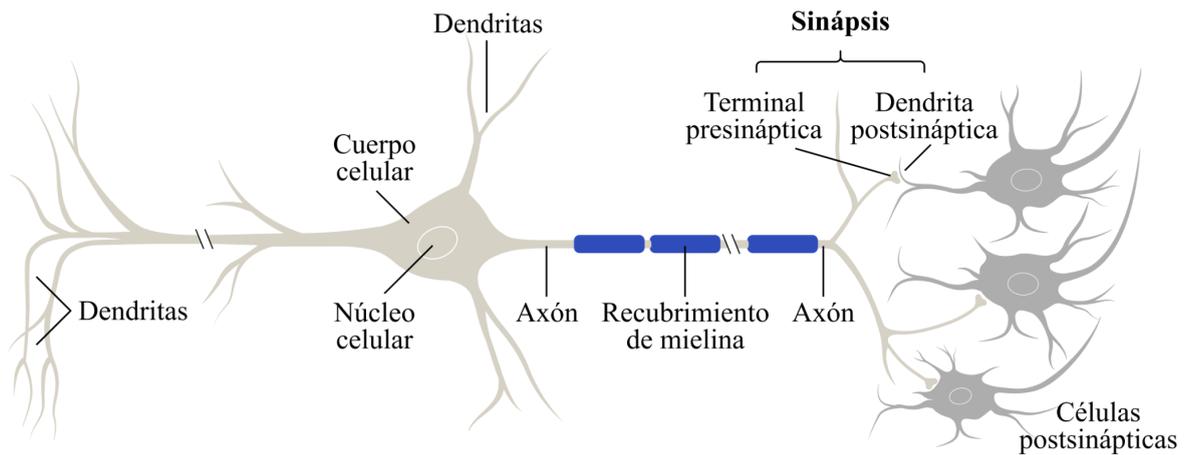
**El axón** empieza en una región abultada del soma que se va alargando hasta dar inicio al cuerpo de axón, este puede extenderse desde un milímetro

hasta 1 metro de longitud. Es habitual que el axón se distribuya en ramas que pueden extenderse largas distancias hasta comunicarse con diferentes partes del sistema nervioso o con las dendritas de otras neuronas (Bear, M. F., et al. 2016). El diámetro de un axón puede variar desde 1  $\mu\text{m}$  hasta los 25  $\mu\text{m}$  en humanos, mientras más delgado es un axón, más rápido se propagan las señales eléctricas a través de él, estas señales son conocidas como potenciales de acción (Kandel, E. R., et al. 2013). En los extremos del axón se encuentran las terminales sinápticas, a través de ellas se hace contacto con otras células y se transfieren los potenciales de acción a ellas. Al punto de contacto se le llama sinapsis (Bear, M. F., et al. 2016).

**Un potencial de acción** es un impulso eléctrico del tipo todo-o-nada y es generado en una región especializada del soma, cerca de la región donde se inicia el axón, desde donde es propagado sin errores o distorsión a una velocidad que puede ir desde 1 m/s hasta los 100 m/s. La amplitud del potencial de acción se mantiene constante, alrededor de los 100 mV mientras es propagado a través del axón, esto se logra regenerando el potencial a intervalos de distancia regulares dentro del axón (Kandel, E. R., et al 2013).

**La sinapsis** se compone de la presinapsis y la postsinapsis, la neurona que emite la señal es llamada presináptica, mientras que la que la recibe es llamada postsináptica (Kandel, E. R., et al. 2013). La presinapsis se encuentra en la terminal del axón y particularmente recibe el nombre de terminal presináptica, mientras que la postsinapsis podría ser una dendrita o el soma de otra neurona. Entre la presinapsis y la postsinapsis se encuentra el espacio sináptico. A la transferencia de una señal desde la presinapsis, pasando por el espacio sináptico hasta la postsinapsis se le llama transmisión sináptica (Bear, M. F., et al. 2016). En la sinapsis, la señal fluye en forma de impulsos eléctricos (potenciales de acción) a través del axón hasta llegar a la terminal donde se haya la presinapsis, ahí son convertidos en señales químicas que cruzan el espacio sináptico y hacen contacto con la membrana de la postsinapsis, entonces las

señales químicas son convertidas de nuevo en impulsos eléctricos. La señal química es llamada neurotransmisor y es almacenada en las vesículas sinápticas al interior de la terminal del axón (Bear, M. F., et al. 2016). La cantidad de neurotransmisor liberado es determinada por el número y frecuencia de los potenciales de acción que alcanzan la presinapsis (Kandel, E. R., et al. 2013).

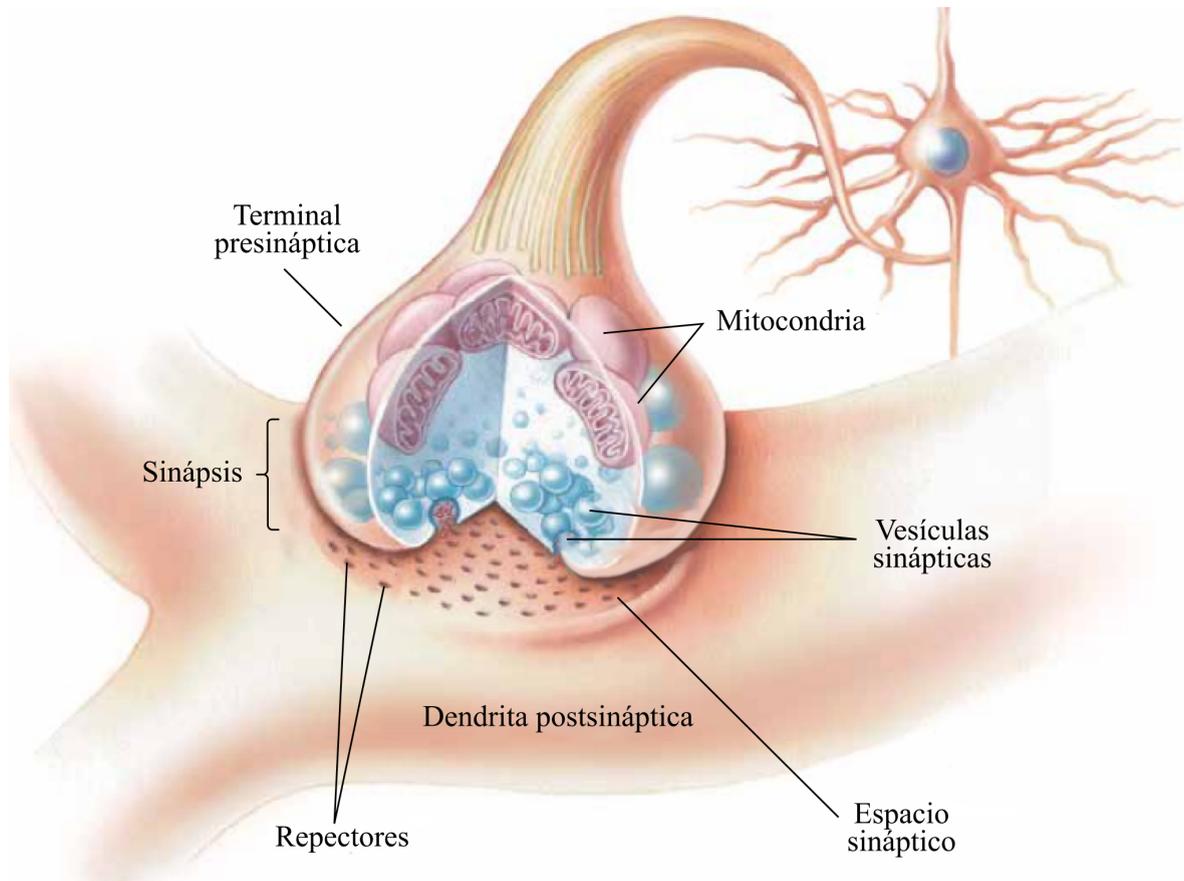


*Figura 2.1.1. Estructura general de una neurona. El cuerpo de la célula contiene el núcleo, donde se almacena el material genético y se originan dos estructuras principales: el axón y las dendritas. El axón suele tener un recubrimiento de mielina. Un potencial de acción se origina en una célula presináptica y propagado a través de su axón hasta alcanzar las dendritas de células postsinápticas mediante la sinapsis. Imagen tomada y modificada de Kandel, E. R., et al. 2013.*

**Las dendritas** son ramificaciones similares a un árbol, estas son el principal mecanismo que tiene una neurona para recibir señales de otras neuronas (Kandel, E. R., et al. 2013), pues operan como la antena de la neurona y suelen estar cubiertas de sinapsis. En la membrana de la postsinapsis existen muchas proteínas, llamadas receptores, encargadas de detectar a un neurotransmisor presente en el espacio sináptico (Bear, M. F., et al. 2016).

En la Figura 2.1.1 se muestra un esquema detallado de la morfología de la neurona y su conexión con otras células nerviosas a través de la sinapsis formando complejas redes neuronales. Mientras que en la Figura 2.1.2 se ilustra

los componentes que forman la sinapsis, entre las que destacan las vesículas sinápticas que liberan los neurotransmisores y los receptores de estos.



*Figura 2.1.2. Terminal del axón formando sinapsis con una dendrita. Cuando un potencial de acción se propaga hasta la terminal presináptica del axón, las vesículas sinápticas liberan neurotransmisores en el espacio sináptico, los cuales se unen a los receptores ubicados en la dendrita postsináptica desencadenando una señal eléctrica en esta. Imagen tomada y modificada de Bear, M. F., et al. 2016.*

Es esta transformación eléctrica-química-eléctrica de señales la que hace posible muchas de las propiedades computacionales del cerebro. Modificar este proceso tiene repercusiones en funciones cognitivas como la memoria y el aprendizaje. Asimismo, disfunciones en la transmisión sináptica apuntan a desórdenes mentales, mientras que muchas de las toxinas y drogas psicoactivas tienen efectos en la sinapsis (Bear, M. F., et al. 2016).

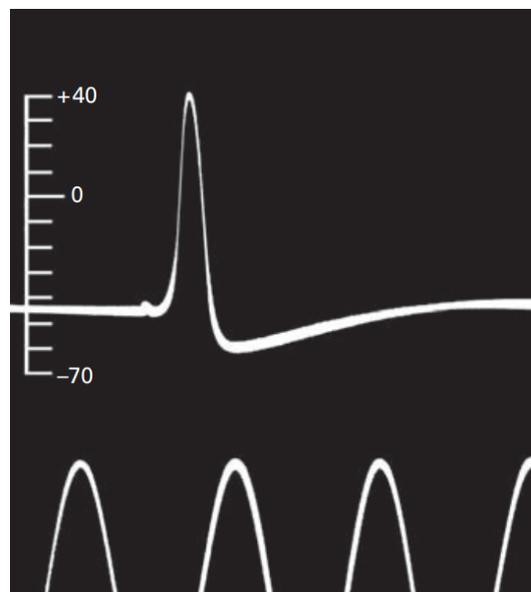
A través de las diferentes características morfológicas de la neurona, se encuentran una amplia variedad de canales de iones ubicados en la membrana celular, los cuales controlan el flujo iones, principalmente  $\text{Na}^+$ ,  $\text{K}^+$ ,  $\text{Ca}^{2+}$  y  $\text{Cl}^-$ , a través de la membrana hacia el interior y exterior de la célula, esto abriendo o cerrándose en respuesta a cambios de voltaje y a señales tanto internas como externas (Dayan, P., et al. 2005).

La diferencia en el potencial eléctrico entre el interior de una neurona y el medio extracelular que le rodea en condiciones de reposo es de alrededor de -70 mV, el medio que rodea la neurona es considerado como 0 mV por convención, en este caso se dice que la célula está polarizada (Dayan, P., et al. 2005). Las bombas de iones ubicadas en la membrana celular son las encargadas de mantener los niveles de concentración de iones adecuados para mantener la diferencia de potencia (Dayan, P., et al. 2005), esto es que el potencial de la membrana en reposo depende de dos factores: una distribución desigual de iones y de la permeabilidad de la membrana (Kandel, E. R., et al 2013). Por ejemplo, la concentración de  $\text{Na}^+$  suele ser mayor afuera de la neurona que al interior, mientras que la concentración de  $\text{K}^+$  es mayor al interior que la exterior (Dayan, P., et al. 2005). Entonces, los iones fluyen al interior o exterior de la célula debido a cambios de voltaje o en la concentración de iones. Una corriente de iones cargada positivamente que fluye hacia afuera de la célula o una cargada negativamente fluyendo hacia su interior, a través de los canales abiertos, hacen que el potencial de la membrana sea más negativo, este proceso es llamado hiperpolarización. Cuando una corriente hace cambiar el potencial de la membrana hacia uno menos negativo o incluso positivo, se trata de un proceso de despolarización (Dayan, P., et al. 2005).

Si una neurona está lo suficientemente despolarizada y se ha incrementado el potencial de la membrana sobre cierto umbral, un proceso de retroalimentación positiva es desencadenado y la neurona genera un potencial de acción, este tiene una fluctuación aproximada de hasta 100 mV que dura

aproximadamente 1 ms (Dayan, P., et al. 2005). Después de haber sido generado un potencial de acción por una neurona, es virtualmente imposible por unos pocos milisegundos que puede generar otro. (Dayan, P., et al. 2005).

Los potenciales de acción son la única forma que tiene una fluctuación en el potencial de la membrana para propagarse largas distancias. Las fluctuaciones por debajo del umbral son atenuadas por completo a distancias no mayores a 1 mm (Dayan, P., et al. 2005). Mientras que los potenciales de acción son regenerados activamente a través del axón que los transporta, haciendo posible alcanzar largas distancias (Dayan, P., et al. 2005). Quizás la característica más sobresaliente de los potenciales de acción es su forma bien definida. Esta característica fue demostrada en la década de 1920 por Edgar Adrian en uno de los primeros estudios del sistema nervioso a nivel celular. Adrian encontró que todos los potenciales de acción tienen la misma forma de onda (Kandel, E. R., et al. 2013) (Figura 2.1.3).



*Figura 2.1.3. Primera publicación de un potencial de acción. Registrado en 1939 por Hodgkin y Huxley.*

En las terminales sinápticas, el voltaje de los potenciales de acción abren los canales de iones produciendo un flujo hacia afuera de la membrana de  $\text{Ca}^{2+}$

seguido por la liberación de un neurotransmisor, este a su vez hace contacto con los receptores ubicados en la postsinapsis, provocando la apertura de sus canales de iones (Dayan, P., et al. 2005). Dependiendo de la naturaleza del flujo de iones, la sinapsis puede ser excitatoria (despolarizando) o inhibitoria (hiperpolarizado) (Dayan, P., et al. 2005).

## **2.2 Imagenología de calcio**

En la investigación sobre el funcionamiento de los circuitos neuronales es importante la habilidad de registrar con precisión y modelar la actividad neuronal de extensas poblaciones de neuronas (Giovannucci, A., et al. 2019). Gracias a los avances en técnicas para la observación y registro de neuronas, tales como imagenología de calcio, se ha vuelto más fácil obtener datos de alta calidad relativos a la actividad neuronal (Li. B. M., et al. 2020); esta técnica se ha posicionado como una de las más importantes herramientas en neurofisiología, ya que permite la observación de actividad neuronal en cientos de células simultáneamente y con resolución de célula única (Kirschbaum, E., et al. 2020) posibilitando el estudio de patrones complejos en la actividad distribuida a través de los circuitos locales de redes neuronales (Ferran, D., et al. 2013).

La imagenología de calcio es un método sensitivo para el monitoreo de actividad neuronal, aprovecha que en células vivas, principalmente las de mamífero, la mayoría de señales eléctricas están asociadas a flujos de calcio  $\text{Ca}^{2+}$  y en particular los canales de  $\text{Ca}^{2+}$  son abundantes en las neuronas (Stosiek, C., et al. 2003). Como es explicado en la sección anterior, los flujos de  $\text{Ca}^{2+}$  forman parte de los mecanismos de comunicación entre neuronas, ya que un potencial de acción al alcanzar las terminales sinápticas provocan la liberación de  $\text{Ca}^{2+}$  y de un neurotransmisor al espacio sináptico, así la presencia de  $\text{Ca}^{2+}$  indica la presencia de potenciales de acción y por lo tanto de actividad neuronal. De esta manera es posible monitorear en tiempo real la actividad a nivel celular,

incluso realizar el registro de múltiples células simultáneamente (Stosiek, C., et al. 2003).

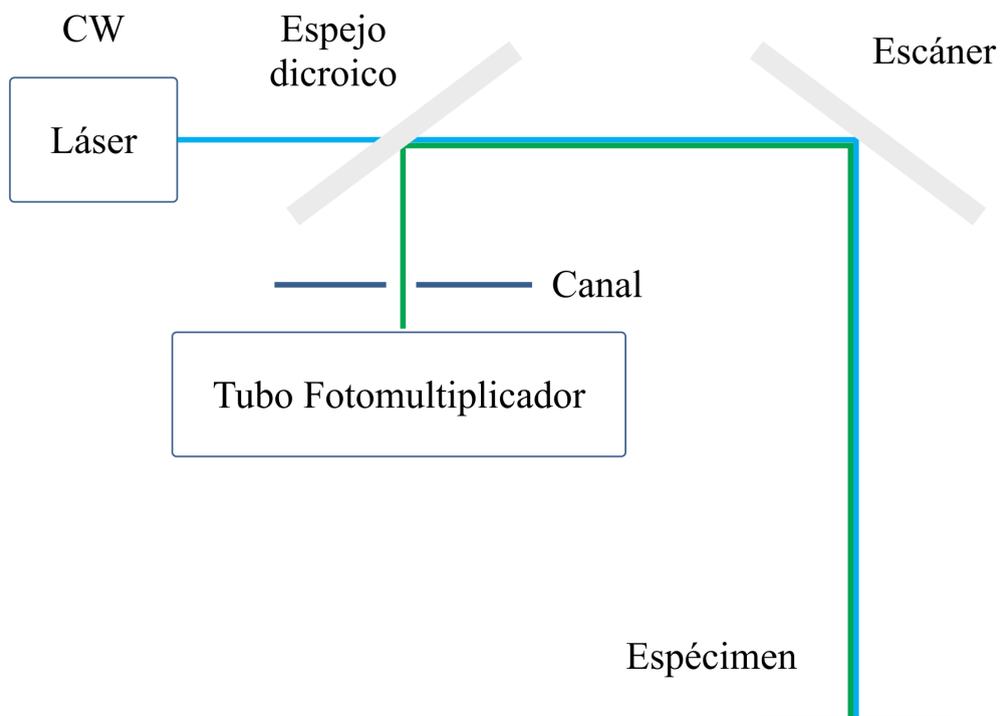
Dentro de la imagenología de calcio, los métodos de observación ópticos basados en fluorescencia de indicadores vinculantes de calcio se han convertido en la herramienta estándar para esta tarea, debido a su habilidad para monitorear células vivas en periodos extensos de tiempo (Giovannucci, A., et al. 2019). En ellos una proteína, el indicador, reacciona con los iones de  $\text{Ca}^{2+}$  produciendo una molécula fluorescente llamada fluoróforo el cual es excitable con luz ultravioleta. Se han desarrollado varios tipos de indicadores de calcio, pero son los genéticamente codificados (*Genetically Encoded Calcium Indicators* GECIs) los que se han vuelto representativos, en especial la familia GCaMP (Grienberger, C., et al. 2012).

La imagenología con resolución de célula única es ampliamente utilizada para el análisis de los mecanismos básicos detrás de la propagación de señales que implican el flujo de  $\text{Ca}^{2+}$  entre neuronas y para el análisis de estructuras neuronales como dendritas y terminales sinápticas. Sin embargo, también es común su uso para monitorear la actividad en poblaciones interconectadas de neuronas (Grienberger, C., et al. 2012).

Es común acoplar un dispositivo de sensado de luz con un microscopio y combinarlo con una fuente de luz apropiada para excitar el fluoróforo liberado por el indicador en presencia de  $\text{Ca}^{2+}$ . La luz empleada para excitar y la emitida por el fluoróforo son separadas por un espejo dicróico ubicado en el microscopio. Entonces la luz es registrada mediante un arreglo de fotodiodos, las cámaras basadas en arreglos del tipo CCD proporcionan alta resolución temporal y espacial, pero suelen ser sensibles al ruido (Grienberger, C., et al. 2012) (Figura 2.2.1). La microscopía basada en escaneo por láser genera la imagen haciendo un barrido con un rayo láser sobre la muestra. Entonces la imagen es creada con los valores de fluorescencia adquiridos para cada píxel por donde pasa el rayo láser. En la microscopía confocal usualmente se emplea

una fuente excitación mono-fotón, así la muestra es iluminada por arriba y debajo de un plano focal (Grienberger, C., et al. 2012).

Para poder aprovechar los datos adquiridos mediante microscopía de epifluorescencia, en forma de vídeo, es necesario extraer de las observaciones la actividad de cada célula de forma individual (Kirschbaum, E., et al. 2020). Sin embargo, la extracción de información es obstaculizada por complicaciones como: i) baja relación señal - ruido, ii) fondo poco uniforme, iii) pérdida de las señales de fluorescencia debido al efecto de fotoblanqueo y iv) traslape de neuronas (Ferran, D., et al. 2013).



*Figura 2.2.1. Instrumentación en microscopía confocal. Se emplea una fuente de luz láser (CW de Continuos Wave), el haz se hace incidir en el espécimen o muestra, la luz reflejada es escaneada y reflejada por un espejo dicroico que separa el haz original del reflejado en un Fotomultiplicador después de pasar por un pequeño canal que filtra una parte de ruido generado por la fluorescencia. Imagen tomada y modificada de Grienberger, C., et al. 2012.*

### 2.3 Experimento característico

Las siguientes descripciones de un experimento característico y del procesamiento de los datos recopilados han sido tomadas de los trabajos de Pérez, J., 2016 y González, F., 2019 respectivamente.

“En el Laboratorio de Neurociencias Cognitivas del IFC-UNAM, se realizan experimentos sobre animales en el marco de las normas mexicanas para la producción, cuidado y uso de animales de laboratorio (NOM-062-ZOO-1999). En un experimento característico, una rata Wistar de 20 a 36 días postnatal es anestesiada con 85 mg/kg de ketamina y 15 mg/kg de xilacina, entonces rápidamente su cerebro es extraído y colocado en una solución salina de 126 mM de NaCl, 3.5 mM de KCl, 26 mM de NaHCO<sub>3</sub>, 1 mM de MgCl<sub>2</sub>, 1.5 mM de CaCl<sub>2</sub>, 1.2 mM de NaH<sub>2</sub>PO<sub>4</sub>, 11 mM de glucosa, 300 mM de mOsm/L, pH=7.4, saturada con 95% O<sub>2</sub> y 5% CO<sub>2</sub> y a 4°C de temperatura. Las rebanadas permanecen en esta solución a temperatura ambiente (21 ~ 25°C) por al menos 1 hora antes de las observaciones. Posteriormente las rebanadas son incubadas en oscuridad de 20 a 25 minutos a temperatura ambiente en presencia de 1 - 10 μM fluo 4-AM (Molecular Probes ®, Life technologies, Ciudad de México) en 0.1% de dimetilsulfóxido (DMSO) y la misma solución salina (Pérez, J., 2016).

Para la observación se emplea un microscopio equipado con objetivo 20X de inmersión en agua y 0.95 NA (XLUMPlanFI, Olympus, Center Valley, PA). Para observar los cambios espontáneos de la fluorescencia se estimula la preparación con pulsos de luz de 488 nm y duración de 15 a 50 ms con una fuente de iluminación Lambda LS (Sutter Instruments, Petaluma, CA) conectada al microscopio mediante fibra óptica. El área de observación consta de 750 x 750 μm. Entonces se adquieren videos mediante una cámara digital con enfriamiento (CoolSnap K4, Photometrics, Tucson, AZ) a una tasa de 100 a 250 ms/cuadro y en intervalos de 5 a 10 minutos. Cada video consta de 3600

cuadros y resolución de 512 X 512 píxeles. En total los videos suman aproximadamente 30 minutos de grabación (Pérez, J., 2016).

Para el procesamiento de los videos, con el software ImageJ los expertos etiquetan las neuronas las regiones de interés ROIs donde se considera que pueden haber neuronas (Figura 2.3.1) y guardan las coordenadas  $(x, y)$  del centro, radio  $r$  y un identificador numérico de cada ROI en un archivo llamado RoiSet.zip con un formato propio de esta herramienta. Posteriormente, mediante la ecuación 2.1 se calcula su señal de cambio de fluorescencia (también llamada transitorio de calcio) para cada ROI, donde  $S_r$  es el valor de la señal de la ROI en un cuadro,  $I_r$  es el promedio de la intensidad de los píxeles que conforman la ROI e  $I_B$  es el promedio de los píxeles contenidos en un área circular con radio  $R$  que rodea la ROI de radio  $r$  y  $R$  es  $k$  veces mayor que  $r$ . Finalmente las señales son procesadas para determinar la ocurrencia de potenciales de acción, aquellas regiones de interés que no presentan actividad son descartadas y los resultados son expresados en una matriz binaria  $N \times F$  donde  $N$  es el número de ROIs,  $F$  el número de cuadros, 0 estado de reposo y 1 actividad asociada a potenciales de acción” (González, F., 2019).

$$S_r = (I_r - I_B) / I_r \quad (\text{Ec. 2.1})$$

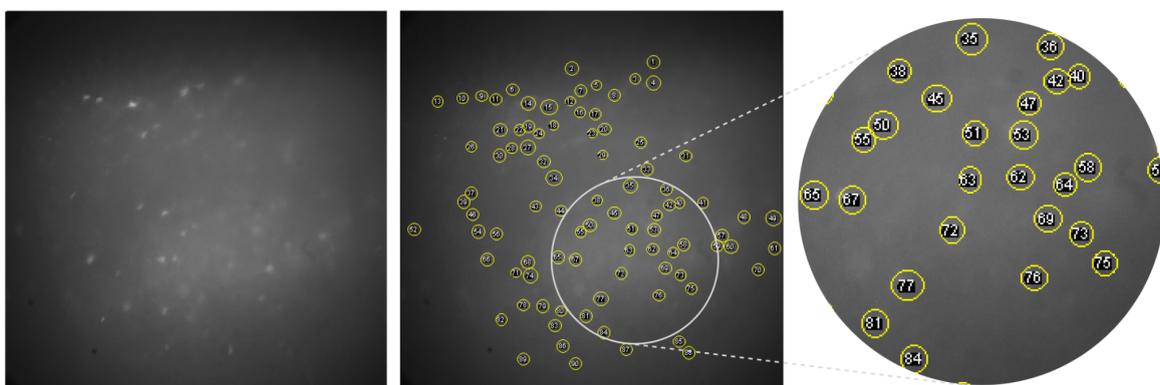


Figura 2.3.1. A la izquierda se muestra una imagen obtenida mediante microscopía de epifluorescencia y a la derecha la misma imagen etiquetada manualmente por un experto.

## 2.4 Redes Neuronales Artificiales

Entre los algoritmos de aprendizaje automático se encuentran dos categorías principales: i) supervisados y ii) no-supervisados. En los algoritmos supervisados, se suministra pares de datos formados por entradas  $x_i$ , también llamadas muestras y su respectiva salida  $y_i$ , también llamada etiqueta, el objetivo es que el algoritmo ajuste parámetros internos con los cuales predecir la etiqueta  $y_i$  para cada muestra  $x_i$ . Si las etiquetas consisten de valores discretos asociados a una clase  $C_k$ , entonces se trata de un problema de clasificación, si son continuos se trata de un problema de regresión. Al proceso de ajuste de parámetros se le llama entrenamiento. Después del entrenamiento, el modelo puede ser utilizado para predecir la salida para entradas nunca antes vistas (Szeliski, R. 2021).

Una de las principales aplicaciones del aprendizaje automático en visión computacional es la segmentación de imágenes, la cuál consiste en clasificar regiones de una imagen. Dentro de las técnicas de aprendizaje profundo, las redes neuronales profundas han sido empleadas con éxito tanto en esta aplicación como en otras, por ejemplo en segmentación semántica, reducción de ruido, estimación de movimiento o detección de objetos (Szeliski, R. 2021).

Las redes neuronales profundas son grafos de cómputo compuestos por cientos de neuronas interconectadas, estas neuronas son unidades de cómputo muy sencillas que realizan, como en una regresión, la suma ponderada de las entradas  $s_i = w_i^T x_i + b_i$ , seguida de una función de activación no-lineal  $y_i = h(s_i)$ , donde  $x_i$  es un vector de entrada para una unidad  $i$ -ésima,  $w_i$  y  $b_i$  son sus parámetros entrenables (también llamados pesos),  $s_i$  es la suma ponderada y  $y_i$  es la salida final de la unidad después de aplicar la función de activación  $h$  a  $s_i$ . Las salidas de cada etapa reciben el nombre de activaciones y se usan como entradas para etapas posteriores (Szeliski, R. 2021).

Las neuronas, en lugar de estar organizadas en grafos irregulares, suelen estructurarse en capas consecutivas, de esta manera se trata a todas las neuronas

contenidas en una capa como un vector, tal que  $s_l = W_l X_l$ , donde  $X_l$  son las entradas a la capa  $l$ ,  $W_l$  es una matriz de parámetros y  $s_l$  es la suma ponderada, entonces, una función de activación  $h$  es aplicada a  $s_l$  tal que  $y_l = h(s_l)$ . Una capa donde la matriz de pesos es completa es conocida como completamente conectada, ya que las entradas a una capa son conectadas a todas las neuronas (Szeliski, R. 2021).

Las funciones de activación son empleadas para propagar las salidas de las neuronas de una capa hacia la siguiente. Estas funciones reciben un escalar y producen otro, se emplean para introducir no-linealidad dentro de las capacidades predictivas de la red y cuantificar cuán cerca está el modelo del estado ideal respecto a las muestras de entrenamiento, para ello se calcula una métrica basada en el error observado en las predicciones de la red. Entonces, se acumula el error de todas las muestras del conjunto de datos y se promedian para obtener un escalar representativo de cuán cerca la red está de estado ideal. La búsqueda de este estado es equivalente a encontrar los parámetros que minimicen la pérdida generada por los errores. De esta manera, es posible enfocar el proceso de entrenar la red como un problema de optimización. En la mayoría de los casos, no es posible calcular estos parámetros analíticamente, sin embargo, pueden ser aproximados mediante algoritmos de optimización iterativos como el descenso del gradiente (Patterson, J., et al. 2017).

Hoy en día existe una gran variedad de funciones de activación, entre las que destaca ReLU (unidad de rectificación lineal), definida como  $h(y) = \max(0, y)$  o la función Sigmoide empleada para realizar regresión logística. En tareas de clasificación, la última capa de la red suele estar destinada a realizar la clasificación final, por lo que una función de activación utilizada generalmente es la función *softmax* que convierte las entradas recibidas por la capa en probabilidades por clase (Szeliski, R. 2021).

El funcionamiento de una red neuronal está dado por su arquitectura, que se define por: i) el número de neuronas, ii) el número de capas, iii) las

conexiones entre estas y iv) funciones de activación. Una vez se ha definido la arquitectura de la red neuronal, inicializando los pesos (comúnmente de forma aleatoria) y la función de pérdida (función que cuantifica el error), es posible iniciar el entrenamiento. Para ello se emplea descenso del gradiente o alguna variante que iterativamente modifican los parámetros, hasta que convergen en un conjunto de valores con una pérdida aceptable en los datos de entrenamiento. Para ello se calculan las derivadas (gradientes) de la función de pérdida por muestra de entrenamiento  $n$  respecto a los pesos  $W$  usando la regla de la cadena, empezando por las salidas y avanzando hacia las entradas. Este proceso es llamado retropropagación (*backpropagation*) y hace referencia a la propagación del error desde la capas finales hacia las iniciales (Patterson, J., et al. 2017, Szeliski, R. 2021). La evaluación y propagación del error a través de la red es realizada capa a capa (Szeliski, R. 2021) y la clave está en distribuir la responsabilidad del error y dividirlo entre los pesos que contribuyeron a la salida (Patterson, J., et al. 2017).

En la mayoría de las muestras de los conjuntos de datos hay ciertas características que están fuertemente relacionadas con ciertas etiquetas. Las redes neuronales aprenden esas relaciones haciendo predicciones basadas en las entradas y los pesos y midiendo cuán exactas son las predicciones. La función de pérdida en algoritmos de optimización como el descenso del gradiente recompensa a la red por buenas predicciones y la penaliza por malas predicciones. Así, los parámetros de la red son movidos hacia las buenas predicciones y lejos de las malas (Patterson, J., et al. 2017). Una de las variantes del descenso del gradiente, es el descenso estocástico del gradiente (SGD), en el cual se divide el conjunto de datos para entrenamiento en subconjuntos aleatoriamente y sin traslape de muestras. Así, en lugar de evaluar todas las muestras de entrenamiento para calcular el valor de pérdida, se realiza con cada muestra aleatoria y se promedia el valor en cada subconjunto (Szeliski, R. 2021).

**Las redes neuronales Convolucionales (CNNs)**, son un tipo de redes neuronales profundas donde las neuronas son organizadas en mapas de características que pueden ser vistas como planos paralelos o canales. Para imágenes, en una capa convolucional, la suma ponderada es realizada empleando una pequeña ventana de la imagen de entrada y los pesos son asociados a los píxeles de la ventana, tal como en la operación convolución. En las redes convolucionales comúnmente se combinan las activaciones de cada uno de los canales de la anterior capa y se aplican *kernels* de convolución por cada uno de los canales de salida. La principal función de las capas convolucionales es la extracción de características locales de una imagen (Szeliski, R. 2021). Las CNN son buenas extrayendo características invariantes a la posición y rotación de los datos crudos de la imagen (Patterson, J., et al. 2017), por lo que este tipo de redes son adecuadas para el reconocimiento de objetos en imágenes, siendo capaces de identificar caras, personas, calles, señalamientos y muchas otras entidades visuales.

**Las redes neuronales recurrentes (RNNs)** son diferentes a otras por su habilidad de procesar información a través de instantes en el tiempo. Son capaces de realizar procesamiento en paralelo y secuencial y, en principio, puede computar lo mismo que una computadora tradicional. Una RNN toma cada vector de una secuencia de vectores de entrada y lo modela, cada uno a la vez. Esto permite a la red retener un estado mientras modela cada vector de entrada a través del flujo de entrada. El modelar la dimensión temporal es el distintivo de las RNNs y lo logra creando ciclos en la red, de ahí el nombre de recurrente (Patterson, J., et al. 2017).

RNNs tienen ciclos en sus conexiones que les permiten modelar el comportamiento temporal en dominios tales como series de tiempo, lenguaje, audio, vídeo o texto. Los datos en esos dominios están inherentemente ordenados y son sensibles al contexto, es decir que los valores futuros se relacionan con los pasados. Los ciclos conectados posibilitan la

retroalimentación, mediante una matriz extra de parámetros para almacenar las conexiones entre cada instante de tiempo, llamada estado oculto (Patterson, J., et al. 2017).

En cada paso de tiempo, procesando las entradas, las neuronas reciben las entradas a través de las conexiones recurrentes que transfieren las activaciones de las mismas neuronas en el estado oculto previo de la red, y el actual vector de entrada. La salida es computada con base en el estado oculto y la entrada actual. Así, los vectores de entrada previos pueden influenciar la salida actual a través de las conexiones recurrentes (Patterson, J., et al. 2017).

Las redes LSTM son el tipo de RNNs más usadas, los componentes críticos en este tipo de redes son: i) la celda de memoria y ii) las compuertas (de olvido y de entrada); el contenido de la celda de memoria está modulado por las compuertas de entrada y de olvido. La configuración mediante compuertas permite a la información fluir para ser retenida selectivamente a través del tiempo y como consecuencia, también los gradientes pueden fluir a través del tiempo, así las LSTM compensan el problema del desvanecimiento o explosión del gradiente que es muy común en las RNNs (Patterson, J., et al. 2017).

Entrenar una RNN es computacionalmente costoso, la opción tradicional es la retropropagación en el tiempo, el cual es fundamentalmente igual al algoritmo de retropropagación, se aplica regla de la cadena para calcular las derivadas con base en la estructura de las conexiones, con la diferencia de que algunas de esos gradientes fluyen hacia atrás, de instantes más recientes hacia los más antiguos (Patterson, J., et al. 2017).

### 3. ESTADO DEL ARTE

---

Mientras que la imagenología de calcio ha permitido realizar observaciones de actividad neuronal a gran escala con resolución de célula única, el desarrollo de herramientas automatizadas, rápidas y confiables para la segmentación de neuronas ha sido un paso crítico en el análisis de este tipo de observaciones (Soltanian-Zadeh, S., et al. 2019), estas suelen implicar varias subtarefas como corrección de movimiento e identificación de potenciales de acción (Klibisz, A., et al. 2017). Generalmente los datos de entrada de estas herramientas se encuentran en formato de vídeo, por lo que están estructurados en tres dimensiones, alto, ancho y tiempo, asimismo la salida producida es una máscara de segmentación en forma de imagen binaria en donde se identifican los píxeles que pertenecen a una neurona (Klibisz, A., et al. 2017).

Aunque se han realizado varias propuestas para resolver el problema de la detección automática de actividad neuronal en observaciones de imagenología de calcio (Apthorpe, N. J., et al. 2016, Baglietto, S., et al. 2018, Giovannucci, A., et al. 2019, Kirschbaum, E., et al. 2020, Klibisz, A., et al. 2017, Lee, Y., et al. 2020), existe un consenso de que permanece como un problema abierto debido a los retos reportados (Baglietto, S., et al. 2018, Giovannucci, A., et al. 2019, Kirschbaum, E., et al. 2020, Stevenson, I., et al. 2011). Entre los principales retos se encuentran aquellos relacionados con las características de las observaciones: i) fuerte presencia de ruido, ii) bajo contraste, iii) bordes difusos, iv) emisión poco uniforme de fluorescencia, v) densidad y ubicación de las neuronas altamente variable, vi) morfología celular heterogénea, vii) ligeros movimientos del tejido y viii) traslape de células.

La identificación de actividad neuronal se basa en dos principales etapas: i) la identificación de las células observadas y ii) la extracción de las señales de transitorios de calcio a través del tiempo por cada célula identificada. La mayoría de los enfoques se ha concentrado en la primera parte, ya que la

segunda consiste en extraer las señales de los datos crudos empleando la ubicación de las células identificadas (Ferran, D., et al. 2013).

Los sistemas para identificación automática de neuronas propuestos pueden ser agrupados mayormente en cuatro categorías: i) factorización de matrices, ii) agrupamiento, iii) aprendizaje de diccionarios y iv) aprendizaje profundo. La factorización de matrices considera el conjunto de imágenes en un vídeo como una matriz 3D y plantea métodos para procesar y extraer señales de ella. Los modelos de agrupamiento definen métricas de similaridad para agrupar píxeles o dividirlos usando algoritmos de agrupamiento. Los métodos basados en aprendizaje de diccionarios, extraen o generan plantillas de neuronas que usan para encontrar nuevas células por similitud. Finalmente, los métodos basados en aprendizaje profundo aprenden características relevantes de las neuronas a partir de conjuntos de datos etiquetados (Klibisz, A., et al. 2017).

Los enfoques tradicionales de segmentación que usan métodos básicos como umbralización u operados morfológicos no son suficientemente precisos. Por otro lado, los enfoques de aprendizaje, ampliamente usados en segmentación de objetos gracias al éxito de las estrategias de aprendizaje profundo, no han resultado adecuados debido a la cantidad de datos etiquetados que se requiere para entrenar este tipo de modelos. (Baglietto, S., et al. 2018).

Los enfoques con base en aprendizaje supervisado pueden ser entrenados para diferentes estrategias de etiquetado y generalizar el aprendizaje a nuevos conjuntos de datos. Estos métodos se han convertido en los enfoques dominantes por alcanzar altos valores de exactitud en muchos problemas de visión computacional (Apthorpe, N., et al. 2016). Sin embargo, mientras que los métodos basados en aprendizaje son útiles para identificar neuronas activas, no lo son para aquellas con baja actividad, haciéndolos inapropiados para casos de estudio donde se involucran neuronas que podrían tener periodos de inactividad (Apthorpe, N., et al. 2016).

Una práctica común ha consistido en tratar de identificar y segmentar las células observadas en una imagen acumulativa o resumen, la cual contiene en cada píxel la proyección del valor promedio o máximo del mismo píxel en el vídeo a través del tiempo, esto implica la pérdida de la información temporal de los vídeos (Kirschbaum, E., et al. 2020, Soltanian-Zadeh, S., et al. 2019). Como consecuencia, los enfoques que se basan en este tipo de imágenes no son competitivos en aquellos conjuntos de datos que consideran neuronas funcionales (Kirschbaum, E., et al. 2020).

La ausencia de conjuntos de datos públicos debidamente etiquetados es un factor crítico que ha impedido, tanto el desarrollo como la correcta evaluación de modelos basados en técnicas de aprendizaje automático (Soltanian-Zadeh, S., et al. 2019). Por tal motivo, muchos modelos han sido probados con conjuntos de datos muy diversos, haciendo difícil realizar comparaciones objetivas (Klibisz, A., et al. 2017). Uno de los conjuntos de datos públicos más populares es *Neurofinder Challenge*, este se compone de 19 videos de microscopía confocal etiquetados manualmente para entrenamiento, además de otros 9 vídeos con etiquetas no públicas para pruebas. Los videos fueron obtenidos bajo diferentes condiciones y etiquetados bajo diversas estrategias, dándose casos en los que se incluyen células inactivas (Kirschbaum, E., et al. 2020). Klibisz, A., et al., han identificado varios retos importantes a considerar en los conjuntos de datos del *Neurofinder Challenge*: son altamente heterogéneos en apariencia, lo que hace difícil definir características comunes para las neuronas, las imágenes tienen alta variabilidad en brillo, las etiquetas están fuertemente desbalanceadas, con un promedio del 12% de los píxeles etiquetados como neuronas. Finalmente, las preferencias de etiquetado presentan inconsistencias, esto implica que un modelo optimizado para las etiquetas de un primer conjunto de datos, sería inconsistente con un segundo conjunto. Asimismo, Soltanian-Zadeh, S., et al. demuestran la existencia de

errores significativos en las etiquetas del *Neurofinder Challenge*, complicando la implementación de soluciones y su adecuada evaluación.

## 4. METODOLOGÍA

---

### 4.1 Análisis del problema

Del estado del arte es posible identificar varios problemas a los que los autores se han enfrentado y por los que ha sido difícil resolver el problema planteado, respecto a los problemas relacionados con las características de los vídeos se tienen: i) fuerte presencia de ruido, ii) bajo contraste, iii) bordes difusos, iv) emisión poco uniforme de fluorescencia, v) densidad y ubicación de las neuronas altamente variable, vi) morfología celular heterogénea, vii) ligeros movimientos del tejido y viii) traslape de células. De estos retos, la fuerte presencia de ruido, el bajo contraste y los bordes difusos, son características relacionadas con la señal del video y podrían ser compensados mediante técnicas de procesamiento de imágenes digitales, en las que se considera a la imagen como una señal en dos dimensiones. Por otra parte, los problemas de emisión poco uniforme de fluorescencia, densidad y ubicación de las neuronas altamente variable, morfología celular heterogénea y traslape de células están relacionadas con la fisiología de las neuronas. Finalmente, los ligeros movimientos tienen su origen en la instrumentación del experimento, pequeñas vibraciones en los dispositivos podrían traducirse en movimiento en los vídeos, este reto podría ser abordado mediante algoritmos de compensación de movimiento.

Los retos relacionados con la fisiología de la neurona podrían ser abordados mediante estrategias de aprendizaje profundo, las cuáles ofrecen la ventaja de que son capaces de aprender a extraer características de los datos que sean apropiadas para realizar la segmentación, a diferencia de otras estrategias donde esas características deben ser extraídas y conocidas previamente. Sin embargo, entre las propuestas basadas en aprendizaje profundo, el reto recurrente es acerca de los conjuntos de datos, este tipo de enfoques requieren de muchos datos etiquetados, el problema no es solo que no hay suficientes

datos, si no que los que hay no están etiquetados de forma consistente debido a que existe una carencia de protocolos estandarizados de etiquetado.

Otro factor de importancia está relacionado con el costo computacional de procesar los vídeos como matrices tridimensionales, por ello una práctica común ha sido la de resumir los vídeos en una sola imagen que sea representativa, por ejemplo obteniendo el valor promedio o máximo de cada píxel a través del tiempo. Esto implica la pérdida de información temporal, así los cambios de fluorescencia característicos de las neuronas funcionales son ignorados, esto es una clara desventaja cuando se requiere de métodos capaces de identificar patrones tanto en el espacio como en el tiempo. Los cambios de fluorescencia son una firma de las neuronas funcionales, no considerarlas evidentemente llevará a segmentaciones poco exactas.

## **4.2 Propuesta**

El sistema propuesto recibe por entrada el conjunto de vídeos pertenecientes a un experimento y por salida genera un mapa segmentación binaria donde se indican como clase positiva a los píxeles que forman parte de una neurona funcional y una matriz llamada *raster* donde se muestra por instantes de tiempo la actividad de cada neurona detectada. La arquitectura general del sistema se compone de tres etapas principales: i) Preprocesamiento, ii) Segmentación y iii) Postprocesamiento (Figura 4.2.1). En la etapa de Preprocesamiento se busca mejorar la calidad de imagen en los vídeos, cuadro a cuadro se aplican operaciones de procesamiento de imágenes digitales con el objetivo de compensar algunas de las características de los vídeos, como ligeros movimientos, ruido y bajo contraste. En la de Segmentación se analizan los cuadros preprocesados para reconocer patrones tanto espaciales como temporales para identificar regiones de interés (ROIs) que pudieran pertenecer a una neurona funcional, estas ROIs quedan plasmadas en una máscara de segmentación binaria. Finalmente, en la etapa de Postprocesamiento se realiza

el refinamiento de la máscara de segmentación y la validación de las ROIs identificadas para determinar cuáles de ellas presentan cambios de fluorescencia característicos de las neuronas funcionales, aquellas que no cumplen con esto son descartadas; como subproducto de este proceso se genera el *raster*.

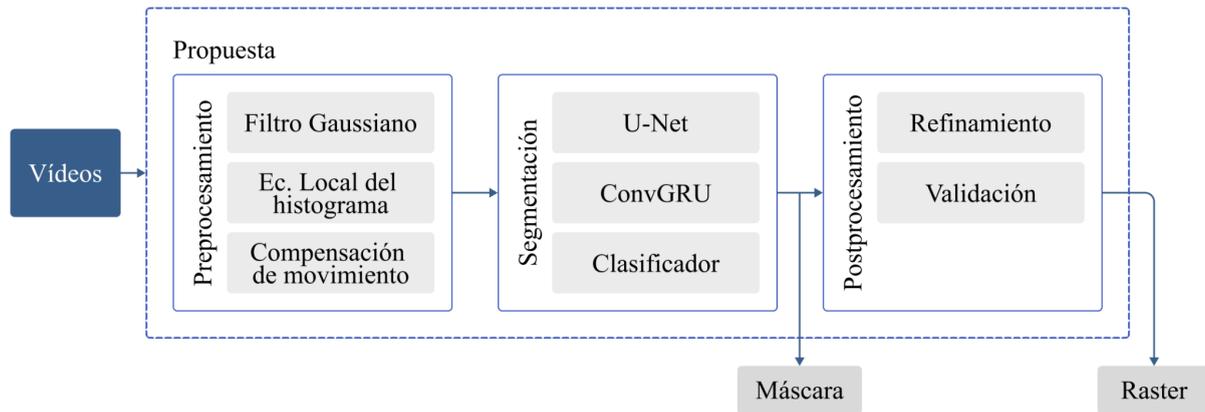


Figura 4.2.1. Arquitectura general de nuestra propuesta.

### 4.3 Preprocesamiento

La mejora de la calidad de imagen de los vídeos consiste en la reducción de ruido mediante la aplicación cuadro a cuadro de un filtro de suavizado Gaussiano de tamaño de 5 x 5 píxeles, en seguida se aplica ecualización local del histograma con ventana de 5 x 5 píxeles para mejorar el contraste. Finalmente, mediante el algoritmo de Lucas-Kanade (Lucas, B. et al. 1981) se calcula el flujo óptico entre dos cuadros consecutivos y los vectores de movimiento obtenidos se utilizan para compensar los desplazamientos.

### 4.4 Segmentación

La identificación de ROIs se realiza mediante dos redes neuronales artificiales (ANNs). La primera es una convolucional tipo U-Net (Ronneberger, O. et al. 2015), la cual integra un bloque *Atrous Spatial Pyramid Pooling* (ASPP) (Chen, L. et al. 2017) como cuello de botella entre el codificador y el decodificador de la U-Net (Figura 4.4.1). La segunda red es una recurrente tipo

*Gated Recurrent Unit* (Siam, M., et al. 2016) que implementa capas convolucionales (ConvGRU) (Figura 4.4.2).

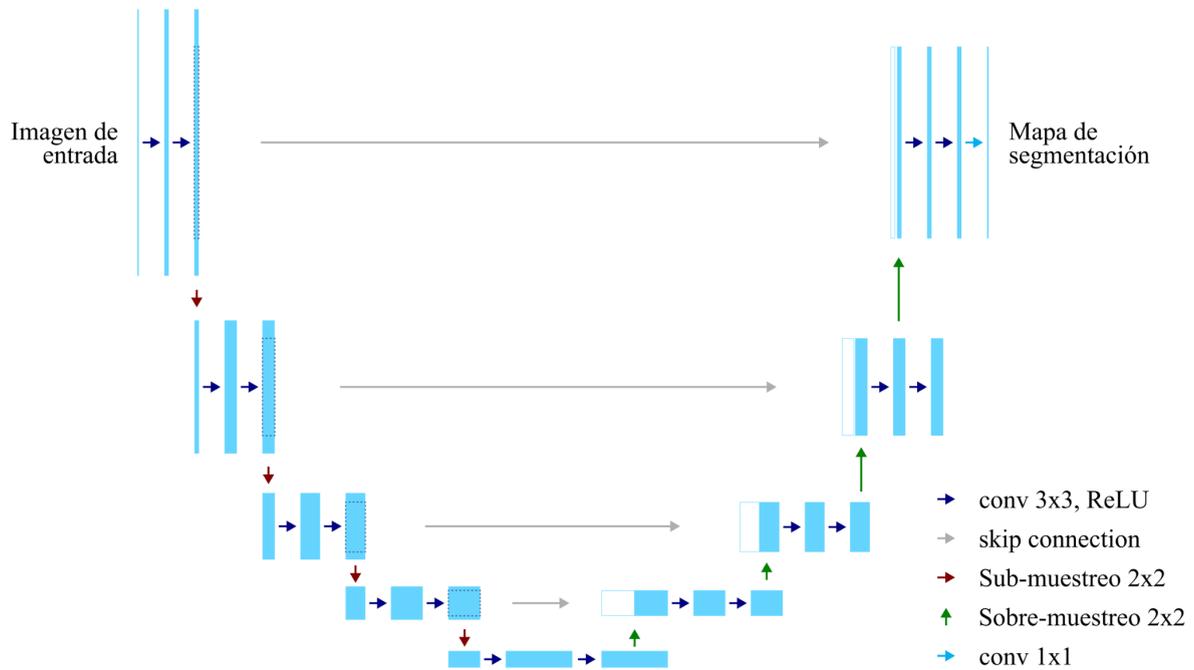


Figura 4.4.1. Arquitectura U-Net, imagen tomada y modificada de Ronneberger, O. et al. 2015.

La U-Net es empleada para la segmentación de pequeñas estructuras en forma de manchas redondeadas (*blobs*) en cada cuadro, estos *blobs* son considerados como potenciales neuronas, cuyo objetivo es identificar todas las células observadas en el cuadro sin importar si son funcionales o no. Entonces la ConvGRU es alimentada de forma secuencial con cada uno de los cuadros segmentados con el objetivo de preservar sólo aquellas regiones con cambios de fluorescencia asociados a potenciales de acción. Al final se agrega una capa convolucional que funciona como un clasificador binario, recibe el último estado oculto de la ConvGRU después de procesar todos los cuadros y produce la máscara de segmentación final (Figura 4.4.3).

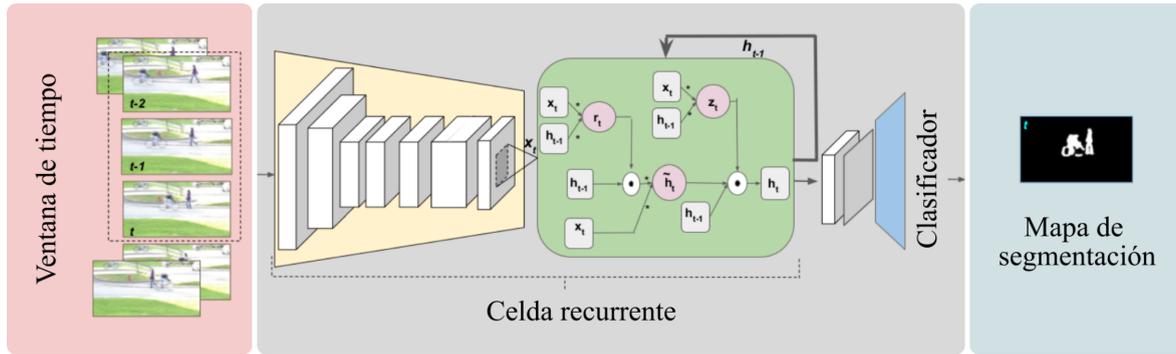


Figura 4.4.2. Arquitectura Conv-GRU, imagen tomada y modificada de Siam, M., et al. 2016.

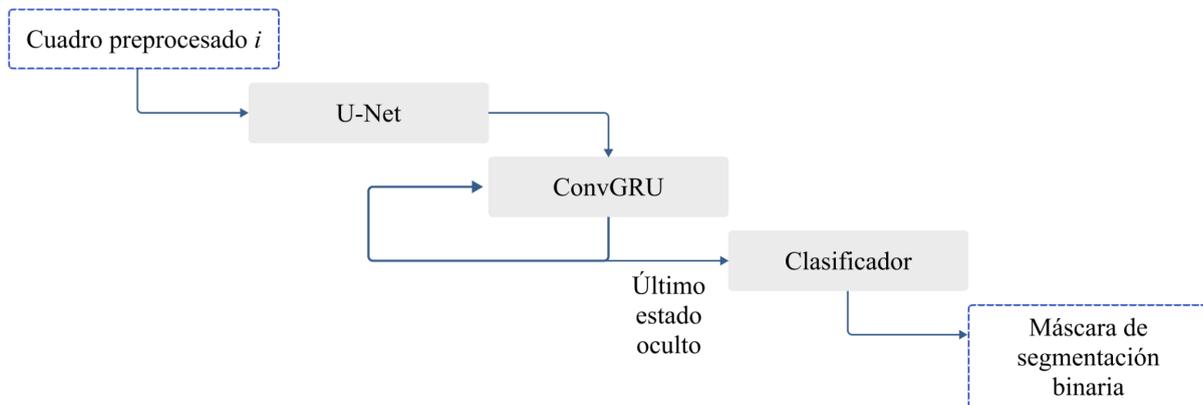


Figura 4.4.3. Flujo de trabajo de la etapa de Segmentación.

El uso combinado de una CNN y una RNN para realizar análisis espacio-temporal como se está proponiendo es un enfoque novedoso para abordar el problema de la detección automática de actividad neuronal en microscopía de epifluorescencia. La arquitectura U-Net ha sido muy popular en la segmentación de imágenes médicas a nivel de píxel (Jha, D. et al. 2020), consiste de una configuración codificador-decodificador, donde el codificador aprende a extraer características y el decodificador aprende a construir la segmentación con base en las características extraídas por el codificador, por esta razón el codificador también es conocido como la etapa de análisis y el decodificador como la etapa de síntesis. Gracias a las *skip connections* que caracterizan a esta arquitectura, es posible propagar los mapas de características generados por las capas convolucionales de la etapa de análisis a las capas

convolucionales de la etapa de síntesis, así no hay pérdida de información y todos los mapas de características pueden ser considerados en la generación de la salida, esto permite la generación de segmentaciones de mejor calidad.

Las RNNs tienen la capacidad de modelar relaciones en secuencias de datos (Day, R. et al. 2017). Sin embargo, las relaciones de largo plazo son problemáticas en las RNNs debido a los problemas del desvanecimiento y explosión del gradiente. Afortunadamente, han sido propuestas arquitecturas que implementan compuertas como la LSTM o la GRU y que resuelven satisfactoriamente estos problemas (Day, R., et al. 2017, Siam, M., et al. 2016). Las arquitecturas con compuertas se basan en modular la entrada actual y el estado anterior de memoria para producir el siguiente estado, debido a que las compuertas tienen sus propios pesos, también pueden ser entrenados para que aprendan a realizar dichas modulaciones mediante el mismo mecanismo del descenso del gradiente (Day, R., et al. 2017). Mientras que la arquitectura LSTM implementa tres diferentes compuertas, la GRU sólo implementa dos, sin embargo, ha sido probado que la GRU presenta un rendimiento similar a la LSTM pero con menos parámetros entrenables (Chung, J., et al. 2014).

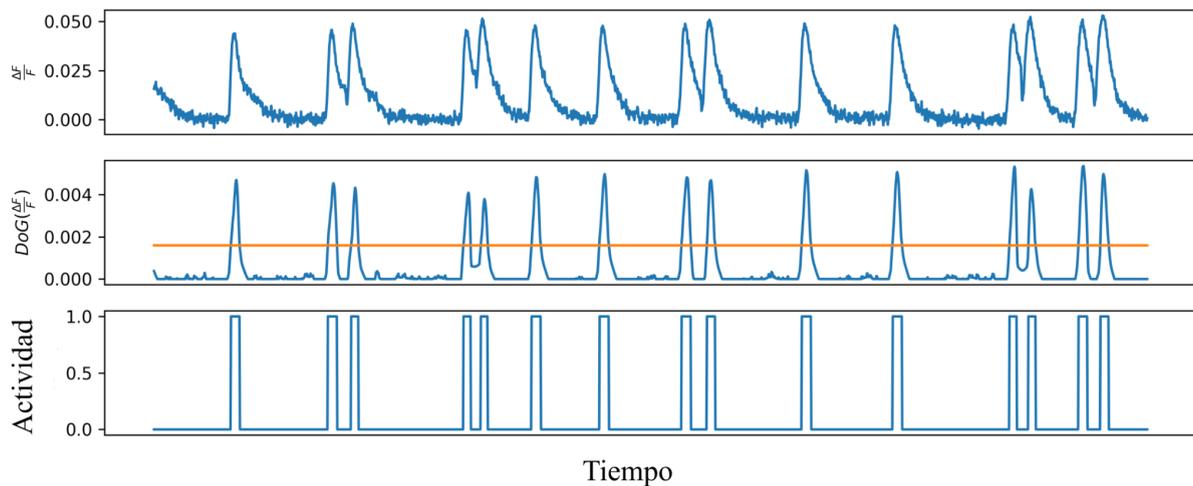
Las RNNs con compuertas fueron diseñadas para procesar vectores unidimensionales, no imágenes o tensores de mayor dimensión. Asimismo, vectorizar una imagen implica dos problemas: i) el vector resultante es muy largo y ii) las relaciones espaciales entre píxeles son ignoradas (Chung, J., et al. 2014). Una solución factible consiste en implementar capas convolucionales en las compuertas, reemplazando el producto punto por un operador convolucional, así el objetivo de las compuertas es aprender los pesos de los filtros en lugar de los pesos por cada píxel (Chung, J., et al. 2014).

Debido a que el problema que se intenta resolver es uno de clasificación binaria, se agrega un clasificador al final de este flujo de trabajo para construir una máscara de segmentación a partir del último estado oculto de la ConvGRU, el cual contiene para cada píxel la probabilidad de pertenecer a una neurona

funcional. Este clasificador consiste de una capa convolucional, normalización por lotes y función de activación Sigmoide.

#### 4.5 Postprocesamiento

La máscara de segmentación generada por la etapa de Segmentación es refinada mediante la eliminación de regiones que son muy pequeñas para ser consideradas ROIs, para ello se aplica un operador de morfología matemática para cerrar las regiones segmentadas. Para validar que la señal de fluorescencia o transitorio de calcio de cada ROI corresponde con el de una neurona funcional, se aplican diferencias de Gaussianas (DoG) para vectores unidimensionales a la señal de fluorescencia y el resultado es binarizado con un umbral establecido en la media más dos veces la desviación estándar, donde la señal supera ese umbral se considera que la neurona está activa (Figura 4.5.1). Las secuencias de actividad de cada neurona son organizadas en una matriz llamada *Raster*, donde cada fila corresponde a una neurona y cada columna a un cuadro.



*Figura 4.5.1. La primera gráfica muestra un transitorio de calcio original de una neurona funcional, en el segunda se muestra el resultado de aplicar DoGs, la línea horizontal corresponde al umbral establecido para binarizar la señal. En la tercera gráfica se muestra el resultado final, los valores en 1 indican instantes de actividad.*

## 4.6 Adquisición de datos

Los datos adquiridos corresponden a experimentos como los descritos en la sección 2.3, de cada experimento se generan entre 3 y 6 vídeos en formato *.avi*, cada uno de ellos contiene 3600 cuadros en escala de grises y con tamaño de 512 x 512 píxeles. Estos vídeos han sido etiquetados de forma manual por expertos. En total se tienen 7 experimentos, equivalente a 30 vídeos, de estos, son reservados 5 experimentos para entrenamiento y dos para pruebas. Los experimentos de entrenamiento suman 24 vídeos mientras que los de pruebas 6.

Los vídeos están contenidos en un directorio por experimento, cada vídeo debe ser primero preprocesado, esto generará un vídeo nuevo por cada uno de ellos. Un conjunto de etiquetas es generado a partir de los vídeos preprocesados; empleando un sistema previamente desarrollado y actualmente en uso en el IFC llamado SRANME (González, F., 2019), se utiliza su módulo de detección de *blobs*, basado en diferencias de Gaussianas, para segmentar todas las potenciales células observadas en 50 cuadros aleatorios de los 30 vídeos. Con este conjunto de  $50 * 30 = 1500$  máscaras de segmentación se genera un conjunto de datos formado por el cuadro preprocesado y la máscara con los *blobs* segmentados, este conjunto servirá para entrenar y evaluar a la red convolucional tipo U-Net (Figura 4.6.1).

Empleando el sistema SRANME, el cual genera una máscara de segmentación con neuronas funcionales a partir de todos los vídeos de un experimento, se generan las máscaras  $y_i$  por cada vídeo original. Asimismo, cada vídeo preprocesado es analizado con la U-Net entrenada, de modo que por cada cuadro se obtiene su respectiva máscara de segmentación con todas las células (*blobs*) contenidas en él, estas máscaras son suministradas como entrada la red ConvGRU y por objetivo se tienen la máscara  $y_i$  generada con SRANME.

En la Tabla 4.6.1 se muestra el conjunto de datos recolectados y su división en conjuntos de entrenamiento y de pruebas.

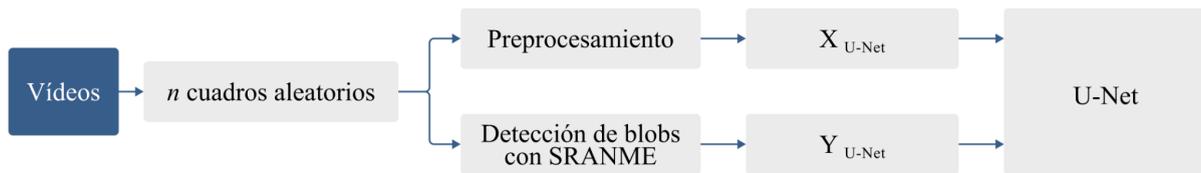


Figura 4.6.1. Entrenamiento de la U-Net.

Tabla 4.6.1. Datos		
Experimento	Núm. Vídeos	Núm. de cuadros aleatorios tomados por vídeo para entrenar la U-Net
Entrenamiento de la U-Net y ConvGRU		
20180324-r1	4	$50 \cdot 4 = 200$
20180325-r1	4	$50 \cdot 4 = 200$
20180325-r2	5	$50 \cdot 5 = 250$
20180721-r3	6	$50 \cdot 6 = 300$
20180721-r4	5	$50 \cdot 5 = 250$
Total	24	1200
Pruebas de la U-Net y ConvGRU		
20180224-r1	3	$3 \cdot 50 = 150$
20180224-r2	3	$3 \cdot 50 = 150$
Total	6	300
Núm. Cuadros por vídeo = 3600 Tamaño del cuadro = 512 x 512 píxeles		

## 4.7 Experimentos

Para entrenar la U-Net, las muestras tomadas de los experimentos para entrenamiento a su vez son divididas en 80% para entrenamiento y 20% para validación, siendo 900 y 300 muestras respectivamente. De los experimentos para pruebas se tomaron 300 muestras. Se empleó aumento de datos mediante la aplicación aleatoria de espejo horizontal y vertical, así como rotación y desplazamiento aleatorio, esto solo para los datos de entrenamiento. Se utilizó

entropía cruzada binaria como función de pérdida, *Adam* como optimizador con tasa de aprendizaje de  $10^{-4}$ , el entrenamiento implementa una estrategia de paro temprano, después de 10 épocas sin mejora en el conjunto de validación se detiene el proceso, en este caso se detuvo después de 47 épocas.

Para entrenar la ConvGRU y el clasificador final se utilizaron los 24 vídeos de los experimentos de entrenamiento dividiéndolos a su vez en 80% para entrenamiento y 20% para pruebas. De los experimentos de pruebas se tienen 6 vídeos. En este caso se tiene por entrada cada vídeo y por objetivo a la máscara de segmentación  $y_i$  generada con SRANME para el vídeo  $i$ , debido a que cada vídeo contiene 3600 cuadros, no es posible realizar el entrenamiento, el cálculo y retropropagación de los gradientes para todo el vídeo, este se divide en bloques de 200 cuadros sin traslape entre ellos, entonces la ConvGRU es entrenada con cada bloque secuencialmente. El último estado oculto del bloque es procesado mediante el clasificador formado por una capa convolucional para generar la máscara de segmentación, esta es comparada con la máscara objetivo  $y_i$  para calcular el error y los gradientes para su retropropagación y finalmente actualizar los pesos de la red. Para evitar la pérdida de la información temporal entre cada bloque, el estado oculto inicial del bloque  $k$ , es el estado oculto final del bloque  $k - 1$  (Figura 4.7.2).

Al igual que con la U-Net, se utilizó entropía cruzada binaria como función de pérdida, *Adam* como optimizador con tasa de aprendizaje igual a  $10^{-3}$ , se implementa una estrategia de paro temprano, la cual detuvo el entrenamiento después de 70 épocas. La ConvGRU usa un *kernel* de tamaño 3 x 3 y un estado oculto con profundidad de 8 canales.

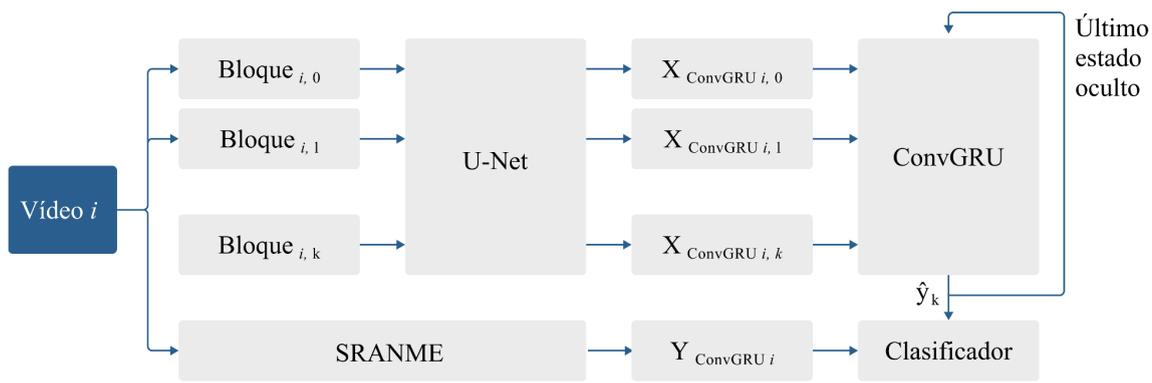


Figura 4.7.2. Entrenamiento de la ConvGRU.

Todas las implementaciones fueron hechas en *python* 3.8, las arquitecturas fueron implementadas a medio nivel en *pytorch* y la ejecución se realizó en una tarjeta de vídeo *Nvidia Titan RTX*.

En el análisis del problema de la sección 4.1 se describe el problema de utilizar imágenes acumulativas, debido a la pérdida de información temporal. En nuestra propuesta se implementa una red ConvGRU para el reconocimiento de patrones temporales, para verificar que su integración ofrece una ventaja sobre las imágenes acumulativas, además de evaluar el rendimiento de nuestra propuesta respecto a la segmentación generada con el sistema SRANME, empleando todas las segmentaciones generadas con la U-Net, se generan dos imágenes acumulativas, calculado el valor máximo y promedio respectivamente; estas imágenes son también comparadas respecto a la máscaras de SRANME.

#### 4.8 Métricas

Para medir el desempeño de nuestras redes se utilizaron las métricas: i) coeficiente de Sørensen-Dice (SDC), ii) Precisión y iii) *Recall*. La primera es adecuada para conjuntos de datos desbalanceados como en este caso, mientras que las últimas dos, si bien pueden estar sesgadas por el desbalance, al compararlas se puede extraer información valiosa como se muestra más adelante. Como se menciona en la sección 1.4, se ha decidido usar el término

*Recall* en inglés debido a que las traducciones de exhaustividad o sensibilidad no transmiten apropiadamente la idea detrás de la relación del número de elementos que un clasificador predijo que era una clase y el número real de elementos que pertenecen a la clase. En la Tabla 4.9.1 se muestran las ecuaciones de las métricas empleadas.

<b>Tabla 4.8.1. Métricas</b>	
Métrica	Ecuación
Sørensen-Dice	$\frac{2 * A \cap B}{ A  +  B }$
Precisión	$\frac{VP}{VP + FP}$
<i>Recall</i>	$\frac{VP}{VP + FN}$
<p><i>VP</i>: Verdaderos Positivos  <i>FP</i>: Falsos Positivos  <i>FN</i>: Falsos Negativos  <i>A</i>: Conjunto de píxeles de la clase <i>A</i>  <i>B</i>: Conjunto de píxeles de la clase <i>B</i></p>	

## 5. RESULTADOS

---

En las Tablas 5.1 y 5.2 se muestran los resultados de evaluar la U-Net y la ConvGRU con sus respectivos conjuntos de pruebas.

En la Figura 5.1 se muestran los resultados de la etapa de preprocesamiento, del detector de *blobs* de SRNAME y de la U-Net, como entrada se ha tomado un cuadro de cada uno de los experimentos de prueba. Visualmente se aprecia que entre ambas máscaras de segmentación hay mucha similitud.

En la Figura 5.2 se muestran, por cada experimento de prueba, las salidas de las máscaras de segmentación producidas por la ConvGRU y resultado después de aplicarles los procesos de Refinamiento y Validación. En este caso se puede observar cómo al final se obtiene una máscara más limpia, sin pequeñas estructuras y las formas de las células se ven más redondeadas y mejor definidas.

En la Tabla 5.3 se muestra el desempeño de nuestra propuesta respecto a las etiquetas manuales hecho por los expertos del IFC, en este caso se ha usado la métrica de *Recall*, la cual nos indica cuántas de las neuronas identificadas por el experto fueron identificadas por nuestra propuesta. En la Figura 5.3 se comparan las células identificadas por nuestro sistema respecto a las identificadas manualmente.

Finalmente, en la Tabla 5.4 se muestra el desempeño de nuestra propuesta respecto al uso de imágenes acumulativas. Se ha usado tanto el promedio como el valor máximo para generar dichas imágenes, esto para cada experimento de prueba. En la Figura 5.4 se muestran las máscaras de segmentación generadas por SRANME, nuestra propuesta y las imágenes acumulativas.

<b>Tabla 5.1. Desempeño de la U-Net en el conjunto de prueba</b>		
SDC	Precisión	<i>Recall</i>
0.814	0.817	0.811

<b>Tabla 5.2. Desempeño de la ConvGRU en el conjunto de prueba</b>		
SDC	Precisión	<i>Recall</i>
0.730	0.721	0.744

<b>Tabla 5.3. <i>Recall</i> de nuestra propuesta en el conjunto de prueba</b>	
Experimento 1	0.7702
Experimento 2	0.7413

<b>Tabla 5.4 Comparativa de nuestra propuesta respecto a las imágenes acumulativas</b>			
Desempeño en el primer experimento			
Método	SDC	Precisión	<i>Recall</i>
Nuestra propuesta	0.7585	0.7452	0.7722
Máximo	0.4392	<b>0.2851</b>	<b>0.9550</b>
Promedio	0.3631	<b>0.7128</b>	<b>0.2436</b>
Desempeño en el segundo experimento			
Método	SDC	Precisión	<i>Recall</i>
Nuestra propuesta	0.7516	0.7061	0.8034
Máximo	0.4529	<b>0.2987</b>	<b>0.9356</b>
Promedio	0.4202	<b>0.7456</b>	<b>0.2925</b>

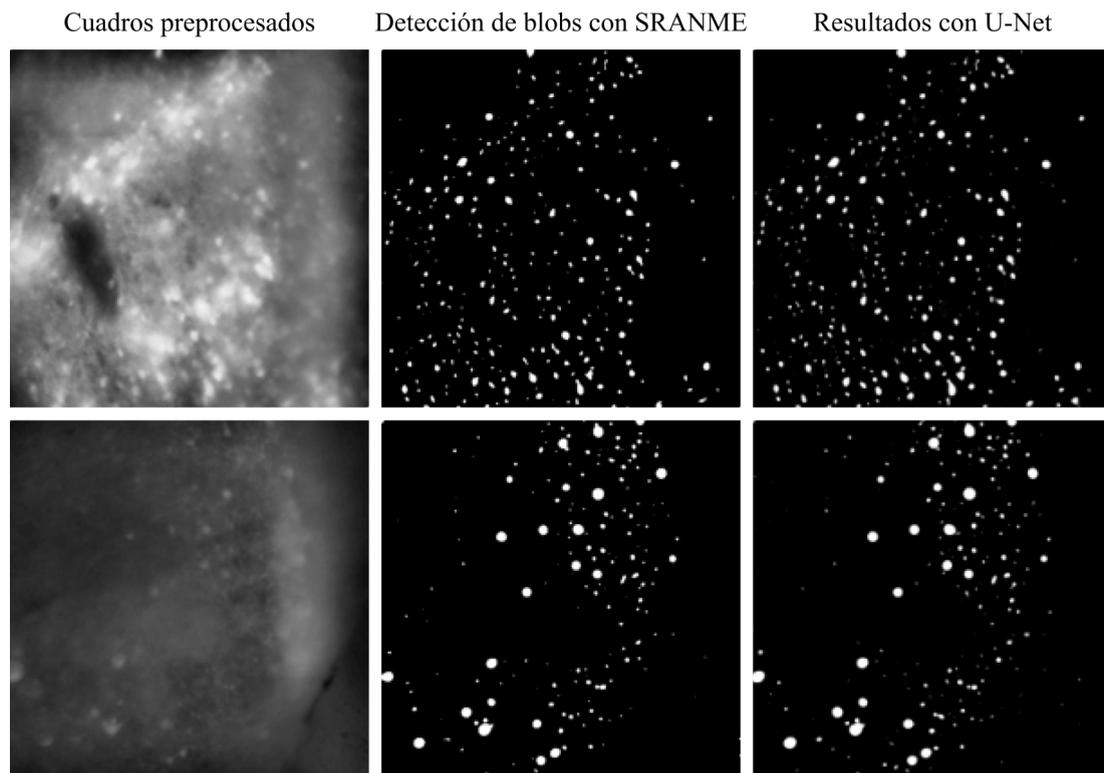


Figura 5.1. Ejemplos de cuadros preprocesados, la máscara generada por el módulo de detección de blobs de SRANME y las máscaras generadas por la U-Net entrenada.

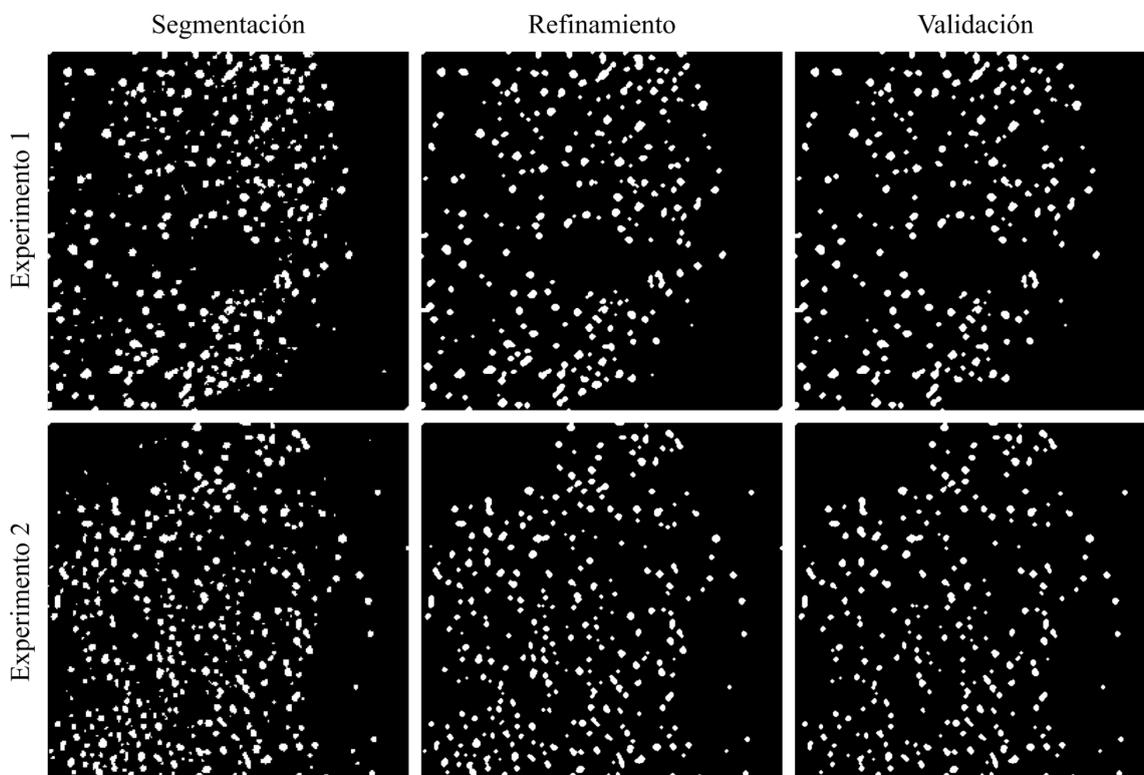


Figura 5.2. Para los dos experimentos de pruebas se muestra la máscara generada por el módulo de Segmentación, el resultado después del refinamiento y de la validación en la etapa de postprocesamiento.

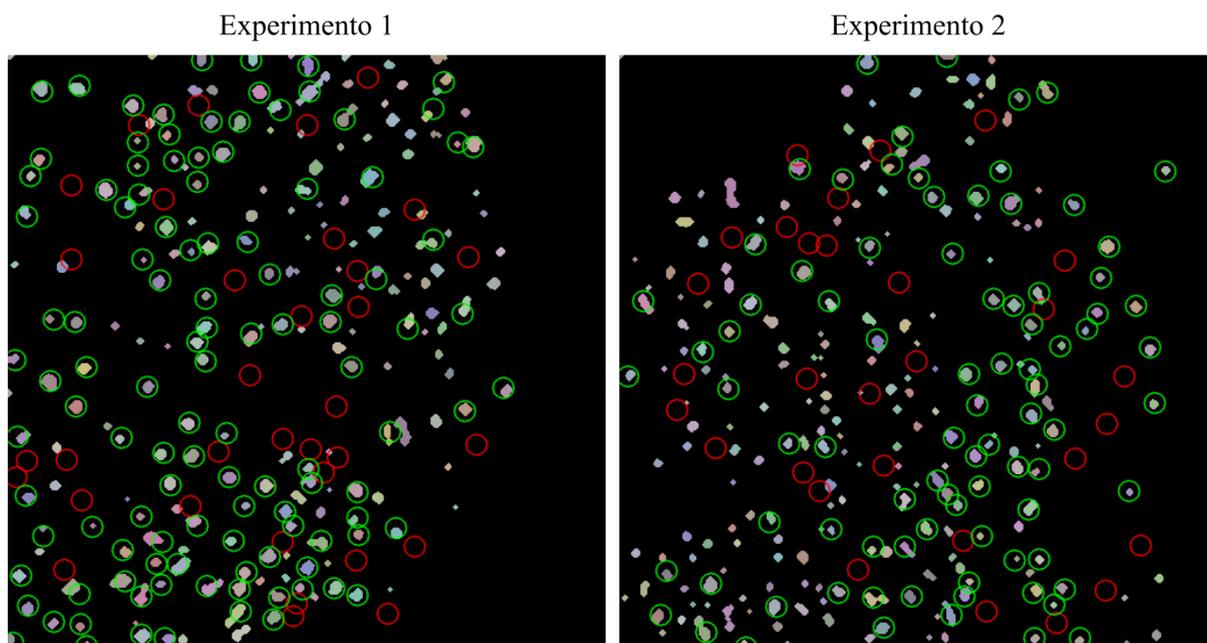


Figura 5.3. Comparación de las máscaras generadas por nuestra propuesta respecto a las etiquetas manuales en los experimentos de prueba. Los círculos verdes corresponden a ROIs identificadas por el sistema y por los expertos, los rojos solo por humanos.

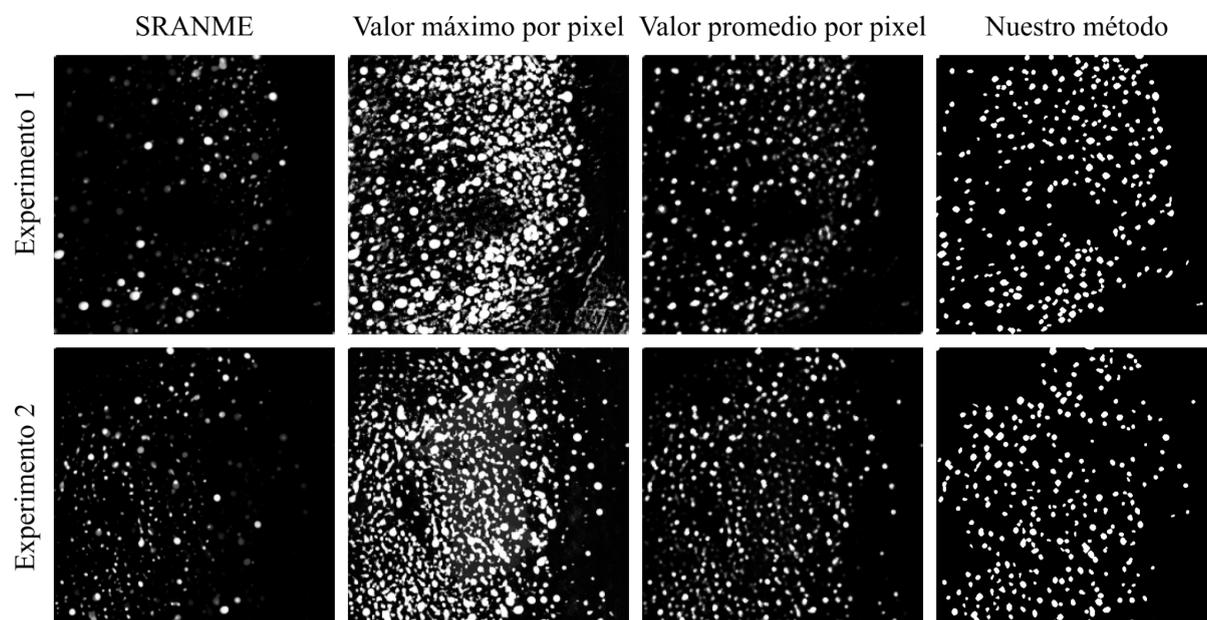


Figura 5.4. Para los dos experimentos de pruebas se muestra la máscara de segmentación generada por SRANME, nuestro método y con él enfoque de imagen acumulativa usando en valor máximo y promedio.

## 5.1 Discusión

En la Tabla 5.4 se observa que en los experimentos de prueba, nuestra propuesta obtiene valores estables en las tres métricas, de  $\sim 0.7$ , a diferencia de los obtenidos usando imágenes acumulativas basadas en el valor máximo y el valor promedio. En estos últimos casos, los resultados de usar una imagen acumulativa con base en el valor máximo son un valor muy bajo en Precisión y uno muy alto en *Recall*, por otro lado, al usar una imagen acumulativa con base en el valor promedio, se tiene el caso contrario, un valor muy alto en Precisión y uno muy bajo en *Recall*.

La métrica de Precisión indica qué cantidad de la región segmentada como positiva realmente lo es. Mientras que la métrica de *Recall* indica qué cantidad de la región que debía ser segmentada como positiva fue segmentada. En este trabajo se ha considerado a la región positiva como aquella que pertenece a una neurona funcional.

Un valor bajo en Precisión y uno alto en *Recall* indican sobre-segmentación, pues la mayor parte de la región positiva es segmentada, pero también lo son muchas regiones que no son positivas, de ahí que en la Figura 5.4 se vean en la imagen acumulativa con base en el valor máximo muchas regiones correspondientes a supuestas neuronas, incluso es imposible apreciar visualmente muchas células individuales pues se encuentran unidas. Por el contrario, un valor alto de Precisión y uno de bajo de *Recall*, indican sub-segmentación, las regiones segmentadas realmente corresponden a regiones positivas, pero sólo se ha segmentado una pequeña parte de la región positiva, en la Figura 5.4 se observa que la muestra correspondiente al valor promedio presenta muchas menos regiones positivas que las demás muestras.

Lo anterior es importante pues demuestra la importancia de tener diferentes métricas que proporcionan información complementaria, de modo que quedarse con sólo una estaría ofreciendo resultados parciales y fáciles de confundir, así mismo, nos permite comprobar numéricamente nuestra premisa

de que el uso de imágenes acumulativas sobre-simplifican la información temporal llevando a segmentaciones de baja calidad, mientras que al considerar cada instante de tiempo se puede aprovechar esa información para obtener mejores segmentaciones.

Al consultar con los expertos sobre los resultados, ellos reconocen que debido a que la tarea de identificar manualmente a las neuronas funcionales es propensa a errores, es muy probable que hayan cometido errores durante el etiquetado, por lo tal motivo se requiere de curar los conjuntos de datos proporcionados para validar qué etiquetas realmente corresponden a neuronas funcionales. Sin embargo, lidiar con aquellas que no fueron etiquetadas es más complejo de resolver. Una solución es un re-etiquetado de las muestras donde los expertos consideren las máscaras de segmentación generadas por nuestra propuesta, para así complementar las etiquetas. Es decir, etiquetar en un proceso iterativo y con ayuda de nuestra propuesta.

## 6. CONCLUSIONES

---

En el presente trabajo se ha descrito el problema de la detección automática de neuronas en observaciones de microscopía de epifluorescencia y la importancia que tiene el poder abordarlo y contribuir en su solución. Se ha dado una breve explicación acerca de los antecedentes de neurociencias para su mejor comprensión. El IFC - UNAM cuenta con investigadores que realizan experimentos con esta técnica de observación y que están interesados en el desarrollo de herramientas automatizadas como la propuesta en el presente trabajo, por lo que han colaborado activamente, proporcionando su conocimiento y datos. Para conocer mejor su labor y el proceso de adquisición de datos, se ha descrito cómo es un experimento típico.

El problema planteado ya ha sido abordado y se han hecho diversas propuestas, incluyendo la realizada por el autor de este trabajo como proyecto de tesis de ingeniería. Después de estudiar el estado del arte y de analizar algunas de las propuestas, se identificaron algunas deficiencias en ellas, como el desaprovechamiento de la información temporal de las observaciones, para compensar y demostrarlo se ha propuesto un sistema que integra técnicas de procesamiento de imágenes digitales, algunas que no están basadas en técnicas de aprendizaje, como filtrado lineal, ecualización del histograma o morfología matemática, más otras que sí, como redes neuronales convolucionales.

La propuesta fue implementada y evaluada con datos etiquetados manualmente por el IFC - UNAM, después de analizar los resultados bajo diferentes métricas, se ha constatado numéricamente, las ventajas que ofrece considerar la información temporal respecto a la simplificación de esta.

Respecto a propuestas que no realizan dicha simplificación, como SRANME, pero que no aprovechan esa información temporal para identificar patrones en el tiempo, también se tienen ventajas, como la capacidad de generalizar el conocimiento adquirido y la posibilidad de realizar transferencia

de aprendizaje para funcionar con otro tipo de experimentos, algo que con sistemas como SRANME implicaría buscar el conjunto de parámetros, que suele ser grande, que mejor se ajuste al nuevo tipo de observaciones.

## **6.1 Trabajo futuro**

Queda mucho trabajo por realizar y cuestiones que resolver, por ejemplo, se describió el problema de la carencia de datos debidamente etiquetados, en ese sentido, el trabajo futuro requiere de recolectar más datos para poder re-entrenar y evaluar nuestra propuesta, así como poder compartirlos como propuesta de *ground truth* para que otras personas hagan sus pruebas y aportaciones.

Es importante dentro de los pasos siguientes, realizar una comparación respecto al estado del arte, optimizar la arquitectura para reducir el costo computacional y con ello el tiempo de espera para los investigadores.

Finalmente, este trabajo se ha basado en experimentos de un tipo de imagenología de calcio, un tipo de proteína fluorescente y un tipo de experimento, a futuro se planea probar con otros tipos de observaciones.

## BIBLIOGRAFÍA

---

Apthorpe, N., Riordan, A., Aguilar, R., Homann, J., Gu, Y., Tank, D. and Seung, H., "Automatic Neuron Detection in Calcium Imaging Data Using Convolutional Networks," 30th Conference on Neural Information Processing Systems (2016).

Bear, M. F., Connors, B. W., and Paradiso, M. A., "Neuroscience: Exploring the Brain," 4th Edition. China: Wolters Kluwer (2016).

Badea, T., Goldberg, J., Mao, B. and Yuste, R., "Calcium Imaging of Epileptiform Events with Single-Cell Resolution", *Journal of Neurobiology* 48, 215-227, (2001).

Baglietto, S., Kepiro, I. E., Hilgen, G., Sernagor, E., Murino, V. and Sona, D., "Automatic Segmentation of Neurons from Fluorescent Microscopy Imaging," *BIOESTEC, CCIS 881*, 121-133 (2018).

Chen, L., Papandreou, G., Schroff, F., and Adam, H., "Rethinking atrous convolution for semantic image segmentation," *arXiv.org* (2017).

Chung, J., Gulcehre, C., Cho, K. and Bengio, Y., "Empirical evaluation of gated recurrent neural networks on sequence modeling," *arXiv.org* (2014).

Cossart, R., Ikegaya, Y. and Yuste, R., "Calcium imaging of cortical networks dynamics". *Cell Calcium* 37, 451–457, (2005).

Day, R. and Salem F. M., "Gate-Variants of Gated Recurrent Unit (GRU) Neural Networks," *IEEE 60th International Midwest Symposium on Circuits and Systems (MWSCAS)*, 1597-1600 (2017).

Dayan, P., and Abbott, L. F., "Theoretical Neuroscience: Computational and Mathematical Modeling of Neural Systems," 1st Edition. London: The MIT Press (2005).

Ferran, D., Reichinnek, S., Both, M. and Hamprecht, F. A., "AUTOMATED IDENTIFICATION OF NEURONAL ACTIVITY FROM CALCIUM IMAGING BY SPARSE DICTIONARY LEARNING," *IEEE 10th International Symposium on Biomedical Imaging* (2013).

Giovannucci, A., Friedrich, J., Gunn, P., Kalfon, J., Brown, B. L., Koay, S. A., Taxidis, J., Najafi, F., Gauthier, J. L., Zhou, P., Khakh, B. S., Tank, D. W.,

Chklovskii, D. B. and Pnevmatikakis, E. A., “CaImAn an open source tool for scalable calcium imaging data analysis,” *eLife Neuroscience* (2019).

González, F., “Sistema de reconocimiento automático de neuronas en microscopía de epifluorescencia,” Tesis de ingeniería, Universidad Nacional Autónoma de México, Ciudad Universitaria, CDMX, (2019).

Grienberger, C., and Konnerth, A., "Imaging Calcium in Neurons," *Neuron* 73, pp. 862-885 (2012).

Jain, A. K., Duin, R. P.W, and Mao, J., “Statistical Pattern Recognition: A review,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Vol. 22, No.1, (2000).

Jha, D., Riegler, M. A., Johansen, D., Halvorsen, P. and Johansen, H. D., “DoubleU-Net: A Deep Convolutional Neural Network for Medical Image Segmentation,” *arXiv.org* (2020).

Kandel, E. R., Schwartz, J. H., Jessell, T. M., Siegelbaum, S. A., and Hudspeth, A. J., “Principles of Neural Science,” 5th Edition. United States: McGraw Hill (2013).

Klibisz, A., Rose, D., Eicholtz, M., Blundon, J. and Zakharenko, S. “Fast, Simple Calcium Imaging Segmentation with Fully Convolutional Networks,” *LMIA/ML-CDS 2017, LNCS 10553*, 285–293 (2017).

Kirschbaum, E., Bailoni, A. and Hamprecht, A. H., “DISCo: Deep Learning, Instance Segmentation, and Correlations for Cell Segmentation in Calcium Imaging,” *ICCAI 2020, LNCS 12265*, 151–162 (2020).

Lee, Y., Xie, J., Lee, E., Sudarsanan, S., Lin, D-T., Chen, R. and Bhattacharyya, S., “Real-Time Neuron Detection and Neural Signal Extraction Platform for Miniature Calcium Imaging,” *Frontiers in Computational Neuroscience Vol 14* (2020).

Li, B. M., Amvrosiadis, T., and Rochefort, N., “CalciumGAN: A Generative Adversarial Network Model for Synthesising Realistic Calcium Imaging Data of Neuronal Populations,” *arXiv.org* (2020).

Lucas, B., and Kanade, T., “An iterative image registration technique with an application to stereo vision,” In *Proceedings of the International Joint Conference on Artificial Intelligence*, 674–679 (1981).

Patterson, J., and Gibson, A., “Deep Learning: A practitioner’s approach,” 1st Edition, O’Reilly Media, Inc. (2017).

Pérez, J., “Características fisiopatológicas de la enfermedad de Parkinson y la discinesia en el conectoma funcional del microcircuito estriatal,” Tesis doctoral. Departamento de Neurociencia Cognitiva, Instituto de Fisiología Celular de la UNAM. Ciudad Universitaria, CDMX, 2016.

Pérez, J., Duhne, M., Lara, E., Plata, V., Gasca, D., Galarraga, E., Hernandez, A. and Bargas, J., "Pathophysiological signatures of functional connectomics in parkinsonian and dyskinetic striatal microcircuit," *Neurobiology of Disease*, 91, 347-61 (2016).

Ronneberger, O., Fischer, P. and Brox, T., “U-Net: Convolutional Networks for Biomedical Image Segmentation,” *arVix.org* (2015).

Serrano-Reyes, M., García-Vilchis, B., Reyes-Chapero, R., Cáceres-Chávez, V. A., Tapia, D., Galarraga, E. and Bargas, J., “Spontaneous Activity of Neuronal Ensembles in Mouse Motor Cortex: Changes after GABAergic Blockade”, *Neuroscience*, (446): 304-322 (2020).

Siam, M., Valipour, S., Jagersand, M. and Ray, N., “Convolutional Gated Recurrent Networks for Video Segmentation,” *arXiv.org* (2016).

Szeliski, R., “Computer Vision: Algorithms and Applications,” 2nd Edition, Springer (2021).

Soltanian-Zadeh, S., Sahingut, K., Blau, S., Gong, Y. and Farsiu, S., “Fast and robust active neuron segmentation in two-photon calcium imaging using spatiotemporal deep learning,” *PNAS* Vol. 116, 8554-8563 (2019).

Stevenson, I. H., and Kording, K. P., “How advances in neural recording affect data analysis,” *Nature Neuroscience*, Vol. 14, 2, 139-142 (2011).

Stosiek, C., Garaschuk, O., Holthoff, K., and Konnerth, A., “In vivo two-photon calcium imaging of neuronal networks,” *PNAS*, Vol. 100, 12, 7319 - 7324 (2013).