

**UNIVERSIDAD AUTÓNOMA DE BAJA CALIFORNIA**

**FACULTAD DE INGENIERÍA, ARQUITECTURA Y DISEÑO**

**MAESTRÍA Y DOCTORADO EN CIENCIAS E INGENIERÍA**



**Uso de métodos de *Deep Learning*  
para clasificación de compuestos  
nanoestructurados heterogéneos**



**TESIS**

que para cubrir parcialmente los requisitos necesarios para obtener

el grado de

**DOCTOR EN CIENCIAS**

Presenta

**CARLOS EDUARDO CABRERA RAMOS**

DIRECTORA DE TESIS

Dra. Dora Luz Flores Gutiérrez

CO-DIRECTORA DE TESIS

Dra. Patricia Juárez Camacho

Ensenada, Baja California, marzo del 2022



# Universidad Autónoma de Baja California

Facultad De Ingeniería, Arquitectura Y Diseño

Uso de métodos de deep learning para clasificación de compuestos  
nanoestructurados heterogéneos

TESIS

que para cubrir parcialmente los requisitos necesarios para obtener el grado de

DOCTOR EN CIENCIAS

presenta:

CARLOS EDUARDO CABRERA RAMOS

Aprobada por:



Dra Dora-Luz Flores Gutiérrez

Directora de tesis



Dra Patricia Juárez Camacho

Co-Directora de tesis



Dr David Cervantes Vásquez

Miembro del comité



Dr Franklin David Muñoz Muñoz

Miembro del comité



Dr Gustavo Alonso Hirata Flores

Miembro del comité

Resumen de la tesis de Carlos Eduardo Cabrera Ramos, como requisito parcial para la obtención del grado de DOCTOR EN CIENCIAS del programa de Maestría y Doctorado en Ciencias e Ingeniería (MYDCI) de la UABC. Ensenada, Baja California, México, marzo del 2022.

## Uso de métodos de *Deep Learning* para clasificación de compuestos nanoestructurados heterogéneos

Resumen Aprobado por:



---

Dra. Dora Luz Flores Gutiérrez  
Directora de Tesis



---

Dra. Patricia Juárez Camacho  
Co-Directora de Tesis

La inteligencia artificial y nanotecnología son dos campos de la ciencia que están cambiando el mundo y empujando hacia novedades en soluciones de la vida diaria. Este trabajo usala inteligencia artificial en su rama de aprendizaje profundo (*Deep Learning*) usando redes neuronales convolucionales en visión computacional mediante imágenes nanoestructuradas heterogéneas obtenidas de microscopía electrónica de transmisión (TEM), a partir de materiales sintetizados que generan luz blanca, para clasificar los compuestos químicos que la integran en las escalas nanométrica y micrométrica. Se desarrolló una metodología que permitió construir 685 imágenes de máscaras binarias (ground truth) a apartir de una colección reducida de imágenes caracterizadas por TEM. Las imágenes están separadas en tres clases diferentes: silicato (silicate), silicón (silicon) y recubrimiento (coating); y cada clase corresponde a los compuestos químicos silicato de itrio (yttrium silicate), óxido de silicio (silicon oxide) y nanopartículas de óxido de silicio fundidas como recubrimiento formando una película delgada (coating) respectivamente. Se usó la arquitectura Mask R-CNN (Region based - Convolutional Neural Networks) para la segmentación de instancias y la arquitectura ResNet 101 como dorsal para la generación de regiones convolucionales, las cuales se entrenaron con las imágenes generadas. El entrenamiento de la arquitectura para clasificación, localización, y segmentación de compuestos químicos logró una precisión por encima del 85% en las tres clases. El modelo generado tiene las capacidades de clasificar nanopartículas superpuestas con una profundidad tridimensional de tres capas de nanoestructuras y aglomerados.

Palabras Clave: Redes Neuronales Convolucionales, Deep Learning, Objetos Traslapados.

## A mis hijos

A esos cuatro que me llaman papá, indefensos y pequeños seres mocosos, juguetones, tragones, gritones, dinámicos, activos, inteligentes, responsables y a veces un poquito pedorritos que hacen mis días llenos de felicidad, que son fantástico regalo cada día para estar con ellos y que son las 4 turbinas de mi vida para trabajar y brindarles un poco de paz en este mundo, pero sobre todo darles todo mi amor.

**Nataly**Alondra, **Carlos**Eduardo, Laura**Carolina** y Richard**Max**Jacob

## A MI ESPOSA

No fue fácil llegar a este momento, ha sido turbulento, cansado, pero sobre todo largo, sin embargo siempre estuviste junto a mi apoyándome con tu esfuerzo y tu paciencia. Te dedico este trabajo porque es un trabajo de los dos, y en todos los momentos de crisis, enfermedad y escasez estuviste a mi lado. Con todo el amor y el cariño de este conjunto de átomos que viven en un entrelazamiento cuántico contigo.

## Dedicatorias

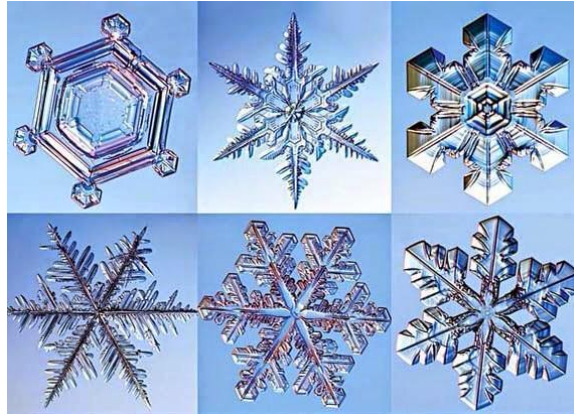
### A mi madre

A esa mujer que a pesar de que los años le han restado energía, fuerzas, movimiento y vigor, sus ojos aún me siguen viendo con la misma intensidad y su corazón amándome con tanta fuerza como cuando me vió por primera vez.

Para ella es todo mi amor, y de regalo este trabajo que hacia muchos pero muchos años se lo había prometido y que pedí a Dios estuviera conmigo para vivirlo juntos.

### A mis hermanas y sobrinos

Las chicas que me conocen de siempre, y han visto toda mi carrera en esta vida, les dedico con todo mi cariño parte de este trabajo porque siempre han sido un apoyo inmedible en mi vida en todo momento, me quedaría corto al decirles que las quiero con todo mi corazón a las tres: Lorena, Denisse y Zulman; y por supuesto no pueden estar fuera de esta ecuación todos los sobrinos que me han dado y a los cuales también adoro.



Un cierto día Dios el Creador discutiendo con Job, le pregunta: ¿Has entrado en los tesoros de la nieve? Me imagino el rostro de Job confundido, sin respuesta, ni siquiera tal vez entendió la pregunta, de aquel compendio de cuestionamientos que se le hicieron.

Si hoy en día nos hiciera esta misma pregunta que le podríamos contestar:

La nieve, como la conocemos, se compone de cristales de hielo. Cada cristal se forma cuando una minúscula gota de agua se congela alrededor de alguna partícula suspendida en la nube y genera un prisma hexagonal, es decir, de seis puntas. Al descender la temperatura de la nube a  $-12^{\circ}$  o  $-13^{\circ}\text{C}$ , las gotas de agua se condensarán y formarán cristales semejantes a ramas de seis puntas. Las mínimas variaciones de temperatura, humedad ambiente, presión, etc., provocarán que cada copo de nieve adquiera una forma distintiva. Esta característica de los copos de nieve los asemeja a las huellas digitales de nuestras manos.

¡No hay dos personas que posean las mismas huellas!

Y si además los vemos bajo el microscopio electrónico veremos tesoros increíbles y hermosas formas.

Gracias a la microscopía electrónica se ha podido revelar el secreto de esta pregunta, y aún existen muchos secretos y misterios escondidos esperando a ser revelados, conforme la ciencia avance y nuevos métodos científicos sean descubiertos en distintas áreas nos seguiremos sorprendiendo de la naturaleza y de la ciencia.

Dedicado a El Creador Dios o la naturaleza



Bosón Cabrera  
(2014-2019)

# Agradecimientos

Quiero expresar mi agradecimiento a la Dra Dora Luz por su dirección durante este doctorado, todo el esfuerzo y energía que invirtió en este período para finalizar el trabajo, y agradecer por su paciencia y el apoyo que me brindó en todo momento y en todas las situaciones. Por el espacio físico y el compartir academia con estudiantes de ingeniería en el laboratorio y el grupo de bio-nano-ingeniería.

A la Dra Patricia Juárez como codirectora por su apoyo total en revisiones y en la integración en su grupo de cáncer y hueso de CICESE, donde tuve la oportunidad de compartir conocimiento con estudiantes de distintas áreas en las ciencias de la vida.

Al Dr David Cervantes por el trabajo de iniciar con la selección de las imágenes que iniciaron como un juego de aprendizaje de IA y terminaron siendo el proyecto principal.

Al Dr Gustavo Hirata por su disponibilidad y apoyo, y además facilitar el contacto que me permitió con sus estudiantes en nanotecnología que compartieron conmigo de sus proyectos y conocimientos.

Al Dr Franklin Muñoz por el excelente curso de nanotecnología que nos impartió y sus observaciones y comentarios acertados al trabajo.

A mis maestros de los cuales aprendí de todo, en áreas nuevas del conocimiento para mí, en cursos de Química con la Dra Eunice, el Dr Angel Fraga con el curso de modelos inteligentes y probabilidad y estadística y por supuesto, aunque fueron de oyente, un excelente curso de Biología Celular y Molecular con el Dr Faustino Camarena y un curso previo de Físico Química con el Dr Alejandro Martínez de la Facultad de Ciencias.

También al MIT y toda esa maquinaria bien coordinada que tienen, donde tuve la aceptación para tomar un curso de Computación Cuántica.

Al grupo de bio-nano-ingeniería, y a los compañeros de laboratorio Carlos Castro y Ernesto León.

Al grupo de cáncer y hueso y todos sus estudiantes, liderado por el Dr Pierrick Fournier y la Dra Juárez por las tantas oportunidades que tuve de perder el miedo de exponer un tema en inglés.

A mis compañeros de la generación y de cursos durante este doctorado.

A los estudiantes de bio-ingeniería Hernan Sandoval y Karla Juárez por su apoyo en el manejo de imágenes, y a Fernanda García por el fuerte apoyo en el trabajo de ground truths y curación

de imágenes.

A los integrantes de nuestras tan amadas reuniones de “*el fin del mundo n-ésima*” Dr Rafael, Dr David, Dr Roberto por su amistad.

A toda la banda de técnicos del Departamento de Oceanografía Física del CICESE.

Al SUTCICESE, por el apoyo en el otorgamiento de la beca para estudios de doctorado al Lic Mario García y al LCC Saúl Rodríguez.

Al CICESE por la beca otorgada apoyada por el exdirector Dr Silvio Guido Marinone y por la beca prórroga apoyada por el director Dr David Covarrubias y el consejo técnico consultivo del CICESE con el apoyo del director de la División de Oceanología el Dr Edgar Pavía y el apoyo de mi jefe inmediato el Dr Rafael Ramírez.

A la Universidad Autónoma de Baja California (UABC) por las facilidades otorgadas para la realización de este trabajo.

A mi laptop ASUS que trabajo durísimo la IA para mi, y al final acabó en manos de los amantes de lo ajeno.

# Contenido

Lista de Figuras	iii
Lista de Tablas	vi
<b>1</b> Introducción	<b>1</b>
1.1 Inicios y evolución en visión artificial .....	1
1.2 Antecedentes en nanoestructuras.....	4
1.3 Descripción del proyecto .....	6
1.4 Planteamiento del problema .....	9
1.5 Marco Teórico .....	10
1.5.1 Nanociencias y nanotecnología .....	10
1.5.2 Inteligencia Artificial.....	10
1.5.2.1 Aprendizaje profundo ( <i>Deep Learning</i> ) .....	11
1.5.2.2 Redes Neuronales Convolucionales.....	13
1.5.2.3 Mask R-CNN .....	15
1.6 Justificación .....	15
1.7 Hipótesis .....	16
1.8 Objetivos .....	16
1.8.1 Objetivo General .....	16

1.8.2	Objetivos Específicos .....	16
<b>2</b>	<b>Metodología</b>	<b>18</b>
2.1	Introducción .....	18
2.2	Imágenes de Microscopía Electrónica de Transmisión .....	18
2.2.0.1	Materiales luminiscentes.....	19
2.3	Creación de <i>Ground Truth's</i> .....	20
2.3.1	Métricas de evaluación .....	22
2.4	Ecualización de imágenes .....	22
2.5	Aumentación de datos .....	25
2.6	Hiperparámetros .....	26
2.7	Entrenamiento .....	27
2.7.1	Cómputo con GPU .....	27
2.7.1.1	TensorFlow.....	28
2.7.1.2	Keras.....	28
2.7.1.3	Lenguaje de programación .....	28
2.7.1.4	Sistemas operativos.....	28
<b>3</b>	<b>Resultados y Discusión</b>	<b>30</b>
3.1	Entropía Cruzada Binaria .....	30
3.2	Entrenamiento .....	31
3.3	Matriz de pesos entrenados.....	33
3.4	Detección, localización, clasificación y segmentación de multiples compuestos .....	33
3.5	Detección y segmentación de compuestos traslapados y superpuestos .....	34
<b>4</b>	<b>Conclusiones y trabajo futuro</b>	<b>39</b>
	Referencias	41

# Lista de Figuras

1-1	Estructura interna de un microscopio de transmisión electrónica (TEM). Tomada y adaptada de Inkson 2016. ....	7
1-2	Esquema de la resolución alcanzada por un microscopio de luz y un microscopio de transmisión de electrones. Tomada y adaptada de <a href="https://ib.bioninja.com.au/">https://ib.bioninja.com.au/</a> . ....	8
1-3	Interacción de los electrones con una muestra en microscopía de transmisión de electrones. (Tomado y adaptado de <a href="https://myscope.training/">https://myscope.training/</a> ).....	9
1-4	Diferencia entre detección, clasificación, segmentación semántica y segmentación por instancias.....	12
1-5	Arquitectura Faster R-CNN.....	14
1-6	Arquitectura Mask R-CNN con sus principales redes: backbone, FPN y Mask. ....	15
2-1	Ejemplo de una imagen TEM seleccionada para el entrenamiento. La imagen muestra compuestos nanoestructurados con material denso de silicato de itrio (área más oscura) .....	19
2-2	Se presentan las tres clases distintas usadas para el entrenamiento. Los recuadros a,b,c son las imágenes TEM, y los recuadros d,e,f son los <i>ground truth</i> creado a partir de las imágenes TEM. a) clase silica, b) clase silicate, y c) clase coating. ....	21
2-3	El recuadro (a) presenta la imagen TEM, sobre la cual observamos a simple vista nanopartículas superpuestas y traslapadas. Se muestran hasta tres niveles de profundidad en nanopartículas de oxido de silicio en los recuadros b,c,d, lo cual genera tres imágenes de ground truth .....	23
2-4	Ajuste del histograma redistribuido $H_{ij}$ (Liu <i>et al.</i> 2019) .....	25
2-5	Diagrama esquemático de una red neuronal profunda ( <i>Deep Learning</i> ). ....	29

3-1	Funciones de pérdida de la arquitectura Mask R-CNN, se observa el desempeño de la red en: a) clasificación, b) segmentación, y c) caja delimitadora. La línea naranja muestra el entrenamiento realizado a 500 pasos por época y la línea azul a 1000 pasos por época. ....	31
3-2	Funciones de pérdida de la arquitectura RPN, se observa el desempeño de la red en: a) clasificación, b) caja delimitadora. La línea naranja muestra el entrenamiento realizado a 500 pasos por época y la línea azul a 1000 pasos por época. ....	32
3-3	Distribución de matriz de pesos: a) Primera capa de convolución de la arquitectura ResNet 101, con un filtro de $7 \times 7$ , con 3 canales y paso igual a 2, generando 64 mapas de características en la convolución. Los percentiles presentan un mínimo en -0.2053, máximo en +0.2185 y una desviación estándar de +0.0385. El histograma representa la distribución de los pesos de los tensores en esta capa. b) Representa la capa de max pooling con un filtro de $3 \times 3$ con paso igual a 2 para reducir la dimensionalidad de la capa beneficiando la invariancia en la traslación. El histograma representa la distribución de pesos de los tensores después del max pooling. ....	34
3-4	Distribución de matriz de pesos en ResNet 101: a) Convolución 2-c con filtro de $1 \times 1$ y 64 canales, creando 256 mapas de características. b) Convolución 3-c con filtro de $1 \times 1$ y 128 canales, creando 512 mapas de características. c) Convolución 4-c con filtro de $1 \times 1$ y 256 canales, creando 1024 mapas de características. d) Convolución 5-c con filtro de $1 \times 1$ y 512 canales, creando 2048 mapas de características.....	35
3-5	Distribución de pesos en Mask R-CNN para: a) predicción de la clase y b) predicción de la máscara. En ambos recuadros se presentan los cuartiles y los histogramas con la distribución de los pesos de los tensores. ....	36
3-6	Distribución de la matriz de pesos de <i>Feature Pyramid Network</i> (FPN) para las capas: a) p5 y b) c5p5. Ambas capas crean una ruta descendente, la cual construye capas de alta resolución a partir de estructuras de alto nivel. ....	36

3-7	Segmentación de imágenes TEM para los tres tipos de clases: silicon (verde), silicate (azul) y coating (cyan). Los recuadros a-c muestran las predicciones de la red neuronal con segmentaciones de la clase silica (verde) con una precisión por encima del 92%. El recuadro c, particularmente muestra la clasificación y localización de las tres clases del experimento (silica, silicate, coating) con el modelo neuronal. Los recuadros d-e muestran la clase silicate inferida correctamente en imágenes a escalas en (d) a 10 nm y en (e) a 50 nm con una precisión por encima del 90%. El recuadro f presenta la segmentación en cyan correspondiente a la clase coating con una precisión por encima del 88%. .....	37
3-8	a) Nanopartículas de silicon detectadas, localizadas y clasificadas en la superficie del silicato de itrio; b) Nanopartículas de óxido de silicio detectadas, localizadas y clasificadas en la superficie del silicato de itrio y superpuestas al silicato de itrio, donde el modelo neuronal tiene la capacidad de detectar dos nanopartículas de silicon sobre la clase de silicate.....	38
3-9	Esta figura muestra las nanopartículas de óxido de silicio (clase silicon) que están traslapadas o superpuestas. a) muestra las nanopartículas segmentadas a mano a simple vista y en b) se muestra la clase silicon como la red neuronal la detectó, localizó, clasificó y segmentó las nanopartículas .....	38

# Lista de Tablas

2.1	Distribución en la base datos de las tres clases de imágenes. ....	19
2.2	Se muestran las tres clases que se identifican en las imagenes por medio de metadatos en formato JSON. ....	20
2.3	Se presentan los métodos de aumentación aplicados a las imágenes aleatoriamente, así como el rango que aplicó el método. ....	25
2.4	Se presentan los hiperparámetros usados durante el entrenamiento de 100 épocas	26
2.5	Tiempo de ejecución de la arquitectura neuronal durante el entrenamiento. ....	27
3.1	Resultados de las funciones de pérdida de las redes Mask R-CNN y RPN después de 100 épocas con 500 y 1000 pasos por época. ....	33

# Capítulo 1

## Introducción

### 1.1 Inicios y evolución en visión artificial

Entendiendo la visión computacional de una forma cotidiana (se llama “visión natural”), se puede comprender como los momentos en que se reúne la familia para ver las fotos de amigos, viajes, una foto antigua, etc.; y tratan de reconocer a todos, y pueden tal vez identificar a las personas por su nombre, parentesco, apodo, si son amigos de la niñez o colegas. Por otro lado, se puede afirmar si están sonriendo, tristes, enojados, dormidos, jugando o bromeando. Si es un paisaje se pueden identificar diferentes escenas del lugar como un río, árboles, cabañas, o si es la foto del equipo de baseball tal vez se identifican uno que otro de los jugadores, o solo a los ídolos del deporte. Definitivamente, de todas las imágenes, siempre hay elementos que se pueden perder por el vago recuerdo, la memoria que falla, por la calidad de la foto, mal tomada, enfocada, borrosa, ya que se pierden detalles y se vuelve complicado al ojo desnudo identificar todos los elementos. Lo importante en la visión natural, es que se pueden hacer todas estas identificaciones y localizaciones de forma ”natural”, banal, rápida ya que el cerebro fue entrenado a lo largo de muchos años con estos recuerdos, entrelazándolos, lo que permite tener esas fracciones de segundo de iluminación donde se trae de la memoria información.

Entonces, la visión computacional (se llama ”visión artificial”), desde la luz de la tecnología, es el arte de observar digitalmente mediante máquinas integradas con sistemas computacionales de visión que identifican y analizan el entorno. Esta disciplina científica, usa métodos y algorit-

mos para analizar e interpretar imágenes del mundo real, produce información práctica y útil del entorno integrada en distintos dispositivos como autos inteligentes, robots, asistentes, para ser usados en la granja, en la oficina, en la ciudad, en la industria y en otras aplicaciones.

Las imágenes en sistemas digitales se generan al colisionar los fotones sobre un panel o celda desplazando los átomos para convertir la luz en electrones, los cuales son leídos para convertirlos en una señal binaria al medir la cantidad de carga en cada celda. Las imágenes en microscopía electrónica de transmisión son generadas con el bombardeo de un haz de electrones sobre una muestra, lo cual permite crear imágenes a muy alta resolución. Al igual que las imágenes adquiridas basadas en fotones, las de microscopía electrónica también tienen parámetros que dependen del usuario como la calidad y nitidez de la imagen, el tiempo de exposición al bombardeo de los electrones, al peso atómico de la muestra, a los ajustes y a la preparación misma de la muestra (Madsen *et al.* 2018). Luego entonces, las imágenes de microscopía electrónica con los escenarios tomados, los materiales y compuestos proyectados se vuelven contextos difíciles de analizar, clasificar e identificar, también, se debe de agregar que el factor de la relación señal-ruido en este tipo de imágenes es bajo.

Resolver este tipo de escenarios hace una década o antes, no solo era una actividad difícil, sino mas bien imposible, principalmente por que los métodos computacionales desarrollados en esta área permanecieron en aplicaciones de bajo desempeño, debido principalmente al bajo poder de cómputo. Sin embargo, los saltos en hardware de alto desempeño computacional en los últimos 10 años, una amplia solución en sensores de alta calidad en muchas áreas, sistemas de almacenamiento incrementados y reducidos en costo (Ley de Kryder<sup>1</sup>), y el abaratamiento y aumento de poder en el hardware (Ley de Moore<sup>2</sup>) y los avances en las tecnologías de procesamiento masivo como las unidades de procesamiento gráfico (GPU por sus siglas en inglés) y la computación de propósito general en unidades de procesamiento gráfico (GPGPU por sus siglas en inglés), han llevado a nuevos límites al desempeño computacional con la gran cantidad de datos que ahora se adquieren y los nuevos métodos computacionales en inteligencia artificial

---

<sup>1</sup>Se basa en el trabajo de Mark Kryder, quien estudió tecnologías de disco duro durante varias décadas, a partir de la década de 1970. Su trabajo inspiró el término ley de Kryder para referirse a los rápidos aumentos en la densidad de almacenamiento de unidades magnéticas durante los últimos 60 años.

<sup>2</sup>La ley de Moore es una observación hecha en 1965 por el cofundador de Intel, Gordon E. Moore, la cual dice que la cantidad de transistores colocados en un circuito integrado o chip se duplica aproximadamente cada dos años. Los expertos creen que es probable que la ley de Moore colapse en los próximos 10 años, debido a eventuales limitaciones materiales.

que requieren de mucho poder de cómputo.

Las implementaciones computacionales de décadas atrás de código inteligente que permanecían en espera de un resurgimiento como el perceptrón (McCulloch y Pitts 1943), el modelo Hebbiano (Hebb y Hebb 1949), las redes neuronales convolucionales (LeCun *et al.* 1998) y el modelo de memoria larga a corto plazo (LSTM por sus siglas en inglés) (Hochreiter y Schmidhuber, 1997) inspirados en los descubrimientos de Hubel y Wiesel, o Q-Learning (Watkins y Dayan 1992), así como los modelos bayesianos, el análisis de componentes principales, entre otros, ahora pasan a ser los primeros actores en la escena manejando grandes volúmenes de datos, a gran velocidad, impulsados por el supercómputo moderno. Pero, además, emergen nuevas estrategias o métodos computacionales en visión computacional basadas en redes neuronales para analizar imágenes como los Vision Transformers (Vaswani *et al.* 2017), que están llamando la atención como la siguiente generación computacional para reemplazar las redes neuronales convolucionales en visión artificial.

Otros desarrollos recientes que han sido importantes en inteligencia artificial son las redes generativas adversarias (GAN por sus siglas en inglés) (Goodfellow *et al.* 2014), las cuales han abierto nuevas estrategias en el área de aumentación de datos en imágenes, al crear nuevos conjuntos sin poder diferenciar entre las imágenes reales y las falsas. Este tipo de redes han creado una amplia variedad de relatos de uso de redes antagonistas.

Como se mencionó en párrafos anteriores, en esta pasada década el hardware para inteligencia artificial ha tenido avances considerables, y han desafiado la capacidad de construir nuevas herramientas y algoritmos en redes neuronales artificiales en entrenamientos distribuidos y paralelos. Esto ha sido trascendental en la parte de optimización en aprendizaje profundo, debido a que el procesamiento por lotes pequeños puede tomar un tiempo considerable suponiendo miles de imágenes de entrenamiento, sin embargo con las nuevas generaciones de tarjetas GPUs, los tiempos de cómputo en comparación cuando inicio la carrera por el aprendizaje profundo era del orden de días, ahora con estos nuevos diseños de hardware con tensores, incluso ya integrado el cálculo de la propagación hacia atrás del perceptrón en cada tensor, puede ser reducido considerablemente el tiempo de cómputo.

De igual manera, los marcos de desarrollo en aprendizaje profundo han evolucionado de una forma permanente ajustándose a los avances en métodos computacionales y al hardware,

estos marcos de desarrollo han permitido diseminar de manera muy importante el uso de la inteligencia artificial en las áreas de aprendizaje de máquina y aprendizaje profundo. Las primeras generaciones como Caffe, Torch y Theano por nombrar algunos, con sus dificultades y limitaciones abrieron el camino a los actuales marcos de desarrollo, versátiles, compactos, muy ambiciosos, fáciles de implementar, además adecuando los algoritmos para el desempeño del hardware en paralelo, como TensorFlow 2.0, PyTorch, Keras, MXNET, los cuales ofrecen distintos niveles de abstracción, son imperativos y ofrecen las facilidades de crear un modelo como conectar piezas de lego.

## 1.2 Antecedentes en nanoestructuras

La segmentación de objetos ocluidos es un problema que se ha estado abordando y tratando de resolver mediante la representación de múltiples objetos ya sean traslapados o superpuestos. Este tipo de escenarios se encuentran en la naturaleza en diversas aplicaciones biológicas como el análisis morfológico de objetos moleculares o celulares en imágenes biomédicas y/o industriales, en el análisis y composición de nanoestructuras o compuestos químicos inorgánicos donde se desea observar cuantitativamente objetos individuales por su tamaño y forma. La interpretación de nanopartículas en resultados experimentales se requiere conocer la distribución de los tamaños de estas nanopartículas (Fisker *et al.* 2000). Por ejemplo las nanopartículas magnéticas se aplican en medios de grabación, ferrofluidos, catalizadores y en biotecnología; sin embargo las nanopartículas magnéticas pequeñas pueden comportarse como superparamagnéticas a temperaturas finitas (Hadjipanayis y Siegel, 1994; Dormann y Fiorani, 1992). Fisker *et al.* 2000 propone un método para determinar el tamaño de las nanopartículas, sin embargo tiene sus limitaciones cuando las nanopartículas están traslapadas o superpuestas.

Otro tipo de soluciones se han propuesto (Kowal *et al.* 2020) para resolver estructuras traslapadas en bioimágenes para segmentar el núcleo basados en umbrales de intensidad, detección de bordes, métodos de transformación de cuencas (*watershed*), contornos activos y redes neuronales artificiales; sin embargo estos métodos suelen tener problemas de identificación por ejemplo con imágenes citológicas debido principalmente a la naturaleza compleja y heterogénea de estas, y que regularmente contienen núcleos superpuestos y en contacto.

En el caso de estudios del cervix, se utilizan técnicas de preparación de la citología para el estudio cervical por los expertos. Esto reduce la presencia de grupos de células y elimina la mayoría de la sangre, el moco y las células inflamatorias en las muestras de citología. Esto permite preparar muestras, que facilitan métodos automáticos para la detección del cáncer cervicouterino; sin embargo, la presencia de residuos, células superficiales y núcleos y citoplasma superpuestos sigue siendo un obstáculo crítico para la detección automática del cáncer de cuello uterino. En Phoulady *et al.* 2016 proponen un método para la detección de regiones donde existen potenciales núcleos en la imagen, con un modelo de mezcla gaussiana enseñado con el algoritmo de *expectation-maximization*, segmentan áreas de agrupamiento de células a partir de intensidades de píxeles, y proponen una métrica para aproximar el límite del citoplasma de cada núcleo detectado en imágenes de múltiples planos focales a través de muestras de frotis de Papanicolaou de base líquida. En Li *et al.* 2016 implementan un modelo de mezcla gaussiana para aplicarlo en cultivos de crecimientos celulares de donde generan miles de videos de clusters celulares, con el fin de investigar e identificar factores de crecimiento celular, división celular o formación de tejido en múltiples condiciones. Las células que crecen en micropatrones tienden a ser compactas, superponerse y traslaparse.

La segmentación semántica de instancias es una técnica en *deep learning* de relevancia en visión computacional para la interpretación de imágenes. El desafío en la segmentación de objetos opacos, translúcidos y transparentes está en distintos dominios como se mencionó anteriormente, y los métodos existentes no pueden resolver esta tarea. En Saleh *et al.* 2019 proponen una arquitectura UNet modificada para segmentar cromosomas traslapados, esto permite dar una opinión a los expertos para el diagnóstico al identificar cromosomas individualmente como el tamaño, índice centromérico, el perfil de forma y los patrones de bandas.

El trabajo de Böhm *et al.* 2019 usa un método de segmentación de instancias semánticas para células que se traslapan y/o tocan, usando la arquitectura UNet. Representan la célula asumiendo el núcleo en el centro y la representan como una tupla  $O_k = (m_k; c_k; b_k; p_k; y_k)$ , donde  $m_k$  es la máscara del objeto,  $c_k$  es el centro del cuadro delimitador,  $y_k$  es la etiqueta de la clase del objeto y  $b_k$  es el cuadro delimitador. Para lograr la detección de estos objetos traslapados primero convierten la escena de 2D a 3D haciendo las capas o niveles separables, permitiendo hacer la predicción de las máscaras sobre estas capas 3D cortadas regresándolas

al espacio 2D. Los problemas encontrados en su desarrollo se encuentran en la segmentación del núcleo, ya que es ligeramente peor que el cuerpo celular debido al desequilibrio de clases y relevante en el caso de objetos pequeños, debido a la separación de las subpartes superpuestas, además que la asignación de los núcleos a sus células es complicado en regiones de alta densidad. En distintas aplicaciones, los objetos frecuentemente se asume que tienen una forma elíptica. Zafari *et al.* 2019 propone un método para detectar nanopartículas basados en la geometría de una elipse, calculando sus semiejes mayor y menor ajustados sobre el contorno de una partícula. Un algoritmo de aprendizaje de máquina no supervisado para detectar nanopartículas metálicas en microscopía TEM desarrollado por Wang *et al.* 2021, tiene la capacidad de reducir el esfuerzo humano en etiquetar los datos. Extrae información morfológica de las nanopartículas metálicas convexas basados en los atributos de sus formas. La forma de las nanopartículas metálicas determinan sus propiedades físicas, químicas y funcionales, el comportamiento plasmónico, la eficiencia catalítica y su actividad biológica. La caracterización en microscopía de transmisión electrónica permite incrementar la eficiencia en la adquisición de datos que actualmente escala el orden de decenas a cientos o más órdenes de magnitud, donde la información excede la capacidad humana de análisis, de aquí que el desarrollo de métodos automáticos de análisis es imperativo. El método requiere de separar el fondo de las nanopartículas y mediante el uso de K-means le agregan la parte no supervisada a la clasificación de las formas que pueden ser segmentadas en las imágenes como solidez, convexidad, área, excentricidad, relación de aspecto y circularidad, usando distribuciones Gaussianas para identificar cada clase.

### 1.3 Descripción del proyecto

La microscopía de transmisión de electrones (Figura 1-1) es una de las técnicas de microscopía más poderosas en la actualidad con magnificaciones de hasta dos millones de veces mayor que la microscopía de luz. La imagen es construida con un haz de electrones que bombardean e impactan sobre una muestra delgada, donde la longitud de onda corta del electrón incidente produce imágenes de alta resolución magnificando partículas de hasta 2 nm.

Las imágenes de microscopía TEM son generadas (Figura 1-3) en tonos de gris o blanco y negro debido a la dispersión de electrones sobre los lentes magnéticos. Si la muestra es densa

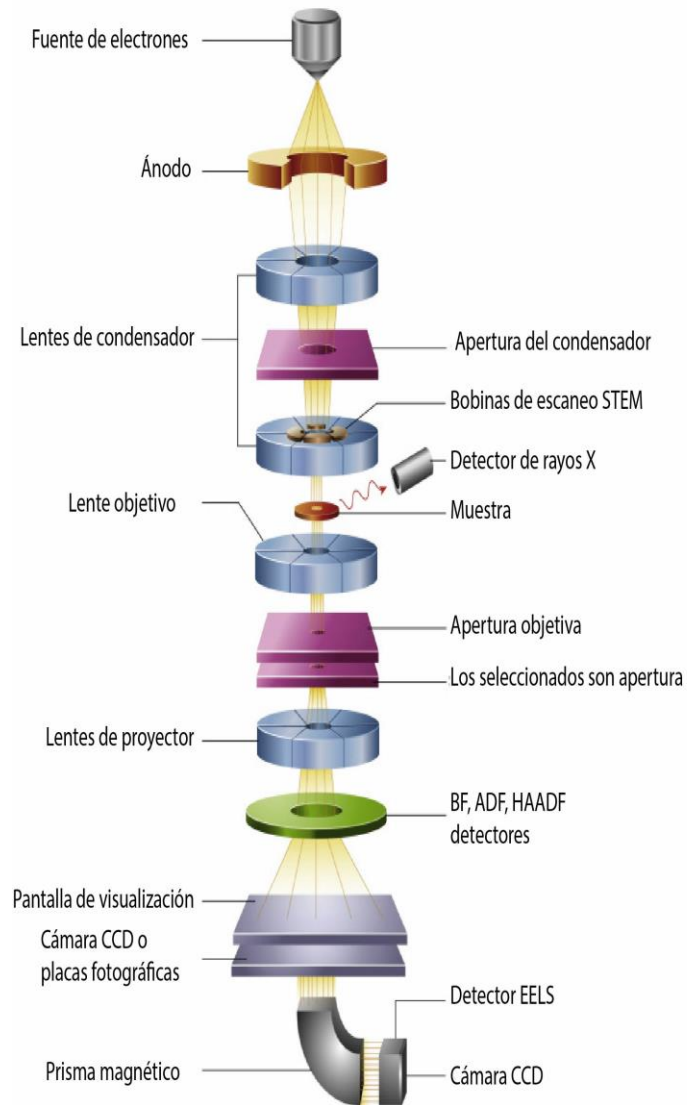


Figura 1-1: Estructura interna de un microscopio de transmisión electrónica (TEM). Tomada y adaptada de Inkson 2016.

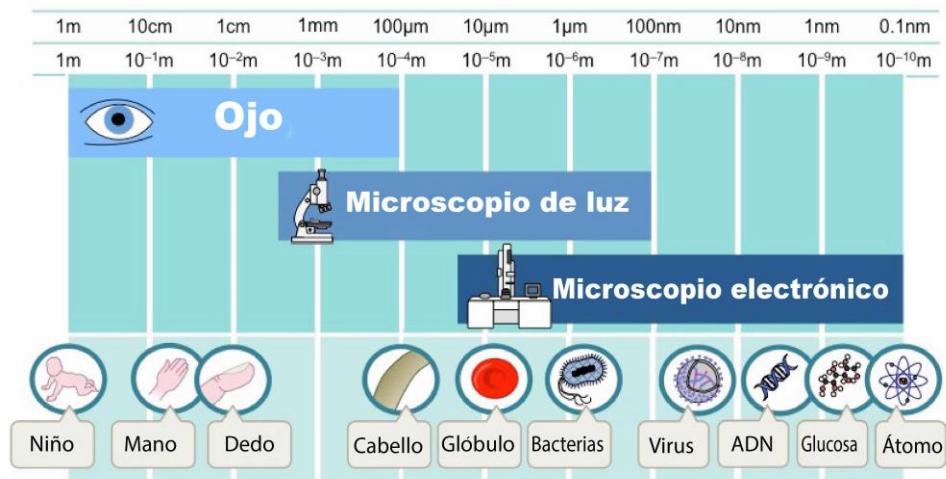


Figura 1-2: Esquema de la resolución alcanzada por un microscopio de luz y un microscopio de transmisión de electrones. Tomada y adaptada de <https://ib.bioninja.com.au/>.

(masa atómica mayor), la cantidad de electrones dispersados incrementa, formando una imagen más oscura, y con una muestra más delgada (masa atómica menor) la imagen aparece más clara o luminosa. Los tonos de gris se vuelven más complejos de discriminar con la cantidad de compuestos que integran la muestra, además el considerar tener un volumen de muestra (escena en 3-dimensiones), también se tienen nanopartículas aglomeradas o nanopartículas ocultas atrás de otras (superpuestas), y/o traslape de estructuras de similar densidad u otra clase de materiales de mayor o menor densidad.

Este proyecto está enfocado al uso de la inteligencia artificial, para localizar, clasificar y segmentar materiales nanoestructurados que un microscopio TEM imprime en imágenes bidimensionales a partir de una muestra (Figura 1-3). Desde el año 2012 se han registrado avances significativos en el uso de la inteligencia artificial en la detección de objetos de distintos escenarios como calles, paisajes, personas, animales, células, construcciones, barcos, y una amplia variedad de objetos de interés (LeCunn et al. 2015; Goodfellow et al. 2016). Las técnicas de aprendizaje profundo (*Deep Learning*) permiten localizar y clasificar objetos en imágenes con un buen desempeño, además son algoritmos rápidos, eficientes y robustos. Los desarrollos e implementaciones actuales están dirigidos a escenarios horizontales, sin profundidad, y regularmente el objeto domina la escena.

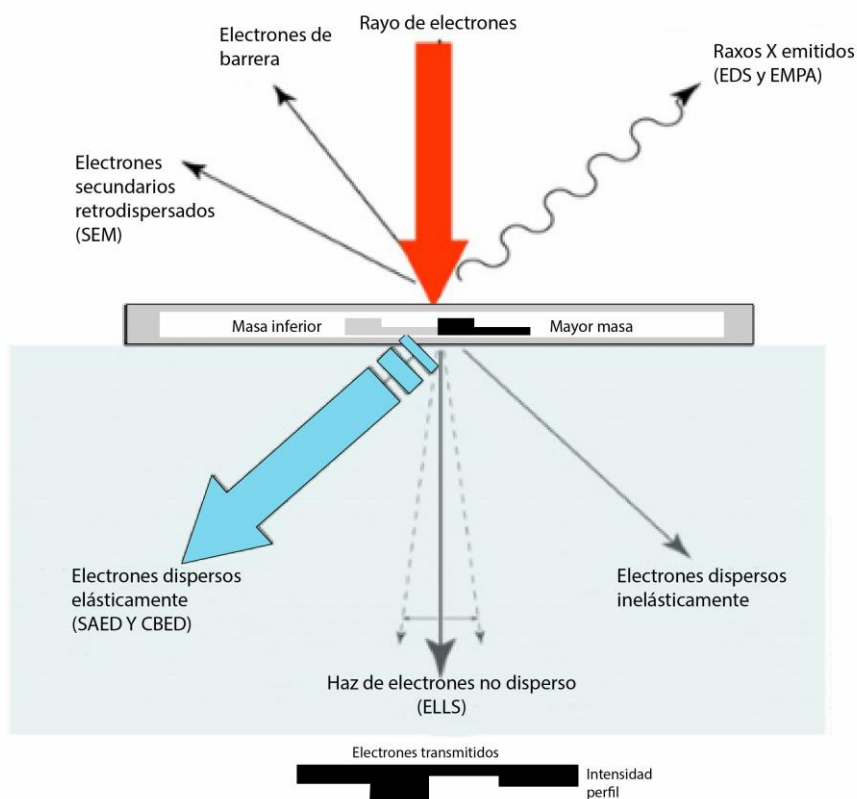


Figura 1-3: Interacción de los electrones con una muestra en microscopía de transmisión de electrones. (Tomado y adaptado de <https://myscope.training>)

Las aplicaciones de inteligencia artificial en las ciencias de los materiales suele ser difícil, ya que estos modelos pueden tener dificultades para aprender características de imágenes significativamente más complejas, con escenarios no tradicionales, y distintos elementos implicando información de tres dimensiones plasmada en un escenario bidimensional. Estos escenarios son tomados de materiales que generan luz blanca (Figura 2-1), y serán analizadas por una arquitectura neuronal para clasificar los elementos y compuestos que integran la imagen.

## 1.4 Planteamiento del problema

En la actualidad existen implementaciones de inteligencia artificial en microscopía electrónica para extraer información estructural atómica (Madsen *et al.* 2018) y en optimización en la

microscopía electrónica de barrido (SEM por sus siglas en inglés) para reducir el tiempo de barrido con la menor pérdida de información (Horwath *et al.* 2020). La microscopía electrónica y la visión computacional son un campo en expansión y requiere de proponer herramientas útiles para los científicos de la microscopía como mejoramiento de imágenes, reconstruir geometrías 3D, y detectar objetos (Ribeiro and Shah 2006). En este sentido, el problema a resolver en este trabajo es el analizar imágenes que contienen nanoestructuras de materiales, donde estos compuestos están dispuestos unos sobre otros, y que a simple vista son objetos difíciles de apreciar, localizar o separar, implicando esto un problema al analizar este tipo de escenarios TEM mediante técnicas de inteligencia artificial, donde se deben detectar y clasificar estas nanoestructuras heterogéneas.

## 1.5 Marco Teórico

### 1.5.1 Nanociencias y nanotecnología

Las nanociencias y la nanotecnología estudian y manipulan la materia respectivamente a escala nanométrica ( $1 \times 10^{-9}$  m). El estudio y manipulación de la materia en sus escalas atómica y molecular está relacionado con las nanociencias (Cervantes *et al.* 2013), donde las propiedades son diferentes significativamente a la de la macroescala. El diseño, síntesis, caracterización, producción y aplicación de estructuras, dispositivos y sistemas en la escala nanométrica está relacionado con la nanotecnología (Bayda *et al.* 2020).

### 1.5.2 Inteligencia Artificial

En la década de los 40's hubo dos grandes acontecimientos que cambiaron la forma de analizar datos: el modelo de la arquitectura de John von Newman (Turing 1936; von Neumann 1947), y la descripción de las redes neuronales artificiales desarrollada con los trabajos de Warren McCulloch y Walter Pitts (McCulloch y Pitts 1943), quienes presentaron un modelo de neurona biológica y los circuitos conceptuales que podrían hacer tareas de cálculo. Aunque, no fue sino hasta finales de la década de los 50's que Frank Rosenblatt introdujo el concepto de perceptrón, el primer modelo neuronal donde la expectativa de esa red neuronal era muy grande, de tal manera que se pensaba que el modelo de ese entonces tendría la capacidad de hablar, pensar,

y caminar, además de otras capacidades.

Los inicios de las redes neuronales fueron basados en el aprendizaje, y no se distinguían por el reconocimiento de patrones, sin embargo eventualmente se fueron desarrollando con mayor complejidad. En los años 70's se desarrollaron los principios de las redes neuronales de "back-propagation", y hasta los años 80's se introdujo la primer red neuronal para reconocimiento de patrones visuales (Fukushima 1980).

Las redes neuronales convolucionales (CNN, por sus siglas en inglés), actualmente usadas en el procesamiento de imágenes, pueden desglosar la información de una imagen en: segmentación de forma, detección de bordes, segmentación de color, segmentación de textura, segmentación de movimiento, detección de objetos, detección de bordes, clasificación de objetos y análisis de la imagen (Affonsoa 2015).

*Deep learning* inicia su auge en el año 2012 con la introducción de la arquitectura AlexNet. Antes de esto, solo se hicieron aplicaciones en el reconocimiento de caracteres, números, e imágenes (Jiuxiang 2017) para aplicaciones en el servicio postal. Los últimos 5 años el aprendizaje profundo se ha estado aplicando en el reconocimiento y análisis de imágenes médicas (Affonsoa 2015; Bienias *et al.* 2019) en radiología, microscopía, tomografía computarizada, resonancia magnética entre otras, donde se obtiene cada vez información más detallada y mas compleja.

#### 1.5.2.1 Aprendizaje profundo (*Deep Learning*)

Los algoritmos o técnicas de *Deep Learning* aprenden niveles de abstracción a través de multiples capas por las que en una dirección (propagación hacia adelante) se les enseña y en la otra dirección (propagacion hacia atrás) aprenden (Figura 2-5). Una gran variedad de algoritmos han sido implementados y mejorados en ciencia de datos, en áreas como reconocimiento de voz, imágenes médicas y detección de objetos, así como aplicaciones en genómica, proteómica, fármacos, etc. (Lecun *et al.* 2015). También han sido usados en áreas de la ciencia y tecnología como robótica, vehículos inteligentes, química, ciencias de la tierra, ciencias del océano, y astronomía.

*Deep learning* usa el enfoque de aprendizaje supervisado, este tipo de entrenamiento, permite que una red neuronal profunda (DNN, por sus siglas en inglés) sea entrenada con datos

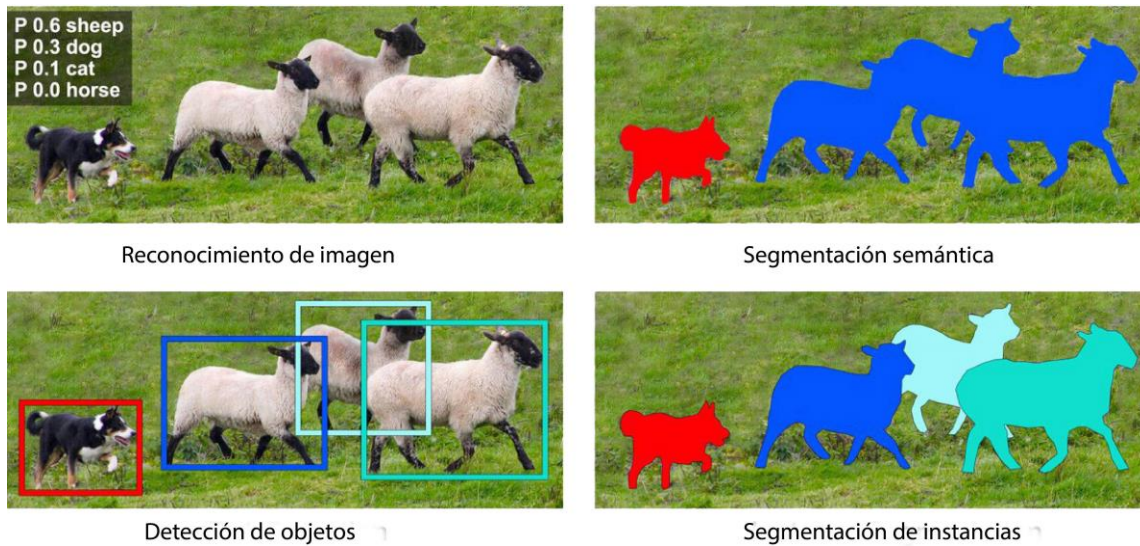


Figura 1-4: Diferencia entre detección, clasificación, segmentación semántica y segmentación por instancias.

etiquetados (Lecun *et al.* 2015). Desde 2015 se han tenido avances significativos en la detección y clasificación de objetos los cuales han sido mejorados usando DNN. Actualmente, la segmentación semántica asocia cada pixel de una imagen con la clase de su etiqueta (carro, casa, perro, gato, célula, etc.). Una técnica reciente, llamada segmentación por instancias, resuelve la compleja tarea de detectar, localizar, clasificar y segmentar cada objeto en un escenario (Figura 1-4). Este método usa la detección de cada objeto en una imagen al nivel de cada pixel, como se hace con el tradicional detector de objetos en visión computacional para localizarlos dentro de una caja delimitada (bounding box). La segmentación por instancias tiene la posibilidad de discriminar entre objetos similares (He *et al.* 2018), el cual es aplicado en los fenotipos de una planta (Romera-Paredes 2016; Yu *et al.* 2020), en vehículos autónomos (De Brabandere *et al.* 2017; Schumann *et al.* 2018), y en la más amplia de las aplicaciones la detección de personas.

Las redes neuronales convolucionales (CNN) reducen el número de hiperparámetros calculados en una red neuronal artificial de manera significativa y son una poderosa herramienta en las técnicas de *Deep Learning* con un rendimiento sobresaliente (Parthasarathy 2017). La reducción de hiperparámetros se debe a que las CNN permiten a sus neuronas conectarse solo

con un subconjunto de la imagen (convolución) en cada capa, en contraste con las redes neuronales artificiales que usan capas totalmente conectadas (Lecun *et al.* 2015). La sinapsis de las neuronas computacionales usan los mismos pesos y se conectan a una pequeña matriz cuadrada impar (convolución) de píxeles del mismo tamaño que la matriz de pesos. Las CNN convolucionan los pesos sobre la imagen completa en cada capa; y en estas mismas capas los pesos son actualizadas durante el proceso de propagación en reversa (aprendizaje) de la red neuronal, generando los nuevos mapas de características entrenados.

### 1.5.2.2 Redes Neuronales Convolucionales

El proyecto ImageNet<sup>3</sup> inició retos provocando el surgimiento de nuevos desarrollos en inteligencia artificial. Este proyecto provee de una base de datos de imágenes muy grande, de fácil acceso, y es el escenario principal para todos los desarrollos de CNN's. Desde que la arquitectura AlexNet ganó el reto de ImageNet en 2012, las CNN se volvieron el estándar de facto en visión computacional. Las CNN son redes neuronales inspiradas en los experimentos desarrollados por David H. Hubel y Torsten Wiesel en el área de la neurofisiología y visión, estudiando las neuronas visuales del cerebro, que están relacionadas con el movimiento, la orientación y el contraste; esta investigación los condujo a recibir el premio Nobel en 1981.

La arquitectura AlexNet es una red neuronal profunda con ocho capas, siendo un trabajo seminal (Romera-Paredes 2016), considerado uno de los de más influencia en visión artificial. Está basado en el modelo LeNet5 de Yann LeCun, el cual a su vez estaba basado en el “NeoCognitron” (Fukushima 1980), la arquitectura original de CNN propuesta por Fukushima. Sin embargo, AlexNet no fue la primera arquitectura CNN, ya que unos años antes Chellapilla en 2006 y Dan Ciresan en 2011 propusieron sus propias arquitecturas. Actualmente, las CNN han mejorado el desempeño visual del humano en algunas actividades.

La arquitectura R-CNN (Girshick 2015; Parthasarathy 2017) se presenta como una red que propone regiones para la detección de objetos, superando el reto de ImageNet de 2013. R-CNN toma una imagen y busca la localización de objetos, cada uno dentro de un cuadro delimitador. Usando el método de búsqueda selectiva sobre las regiones propuestas busca el cuadro delimitador (bounding box). La arquitectura usa la región propuesta, la cual es pasada

---

<sup>3</sup><http://www.image-net.com>



Figura 1-5: Arquitectura Faster R-CNN.

a un tamaño estandar y entonces pasa a la red AlexNet. La última capa de R-CNN agrega la clasificación de los objetos con una máquina de soporte de vectores. El último paso aplica una regresión lineal sobre las propuestas de regiones para generar las coordenadas de los cuadros delimitadores más exactos.

La arquitectura Fast R-CNN (Girshick 2015; Ren 2016) fue construida sobre el trabajo previo de R-CNN, mejorando la respuesta y precisión, agregando algunas innovaciones a la red y se introdujo el concepto de “Region of Interest Pooling (RoIPool)”, el cual reduce los calculos en la arquitectura.

La arquitectura Faster R-CNN (Ren 2016) introdujo “Region Proposal Network (RPN)”, mejorando significativamente la velocidad de respuesta para la propuesta de regiones, usando la capa de convolución en dos vías, una como la propuesta de regiones, y la otra como clasificador. La arquitectura Fast R-CNN habia reducido el tiempo de ejecución en RPN por un cuello de botella, el cual fue mejorado en Faster R-CNN.

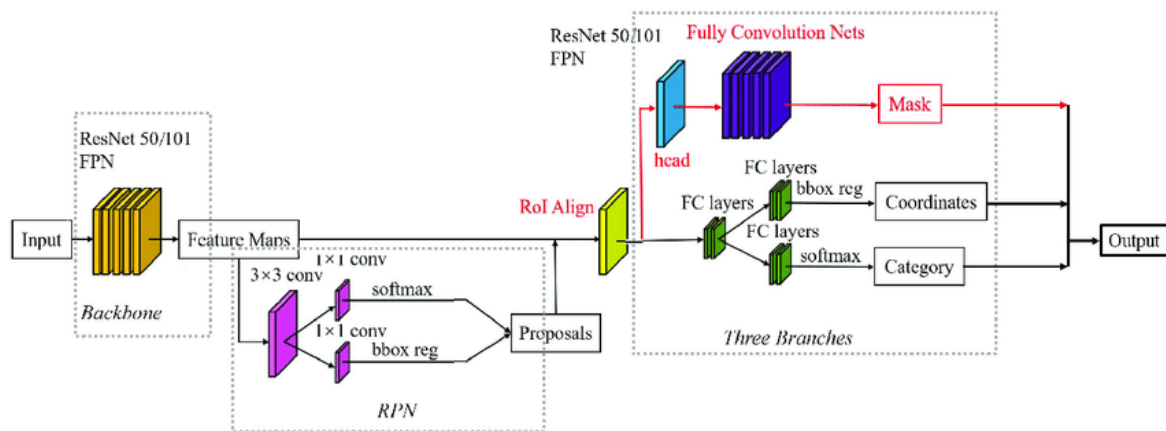


Figura 1-6: Arquitectura Mask R-CNN con sus principales redes: backbone, RPN, FPN y Mask.

### 1.5.2.3 Mask R-CNN

La arquitectura Mask R-CNN (He 2018) está diseñada para desempeñar segmentación de instancias, mejorando la arquitectura anterior Faster R-CNN (Figure 1-6), agregando un brazo (Fully Convolutional Network) con una máscara binaria de salida al nivel de pixel que responde con un 1 si el pixel es parte de un objeto (clase) o un 0 si no. Además, Mask R-CNN predice un puntaje (score), clase, localización de los pixeles exactos sobre cada región de interés (RoI), la caja delimitadora (bounding box) y la máscara (con 0's y 1's) para generar la predicción de las instancias segmentadas.

## 1.6 Justificación

La visión computacional es un campo en expansión en todas las áreas de la ciencia, extendiéndose al campo de la nanotecnología, impactando con desarrollos recientes con considerables esfuerzos en microscopía donde se generan escenarios que implican un reto a la visión computacional debido a las condiciones de baja resolución o imágenes con una baja relación de señal a ruido, por considerar algunas. Esto implica desarrollar nuevas técnicas computacionales, nuevos algoritmos o nuevas metodologías que se adapten al análisis e interpretación de este tipo de imágenes.

El reto es evaluar un escenario tridimensional, proyectado en un ambiente bidimensional.

Los sistemas actuales de detección y clasificación evalúan escenas generalmente donde el objeto domina, no importa la detección del fondo o el ocultamiento de objetos. La principal problemática en este tipo de imágenes de microscopía TEM es el desarrollo de un sistema o metodología de inteligencia artificial con la capacidad de identificar materiales nanoestructurados heterogéneos superpuestos y/o aglomerados en los tres tipos o clases de materiales usados en la generación de luz blanca que capta la microscopía, y por supuesto para los niveles como tantos elementos contenga el escenario.

La importancia de considerar la evaluación de estos escenarios, implica que para este tipo de microscopía electrónica se pueden evaluar los escenarios y los materiales con sus características, elementos, objetos, y clases incluidos. Esta metodología se puede extender a nanotubos de carbono para depósito de capas, o en el caso de imágenes de arreglos atómicos clasificar el(os) átomos en la estructura. En biología, se puede extender a segmentar células y organelos traslapados. En el caso de imágenes de Rx, se puede implementar para determinar estructuras ocultas de tejidos.

## 1.7 Hipótesis

El cuestionamiento en este trabajo es, determinar, si es posible que al implementar las técnicas o metodologías de segmentación de instancias, con los algoritmos computacionales de *Deep Learning* del tipo Mask R-CNN son eficientes y funcionales para segmentar, clasificar, localizar y detectar nanoestructuras heterogéneas aglomeradas, superpuestas, traslapadas y aisladas en imágenes TEM de silicato de itrio y óxido de silicio.

## 1.8 Objetivos

### 1.8.1 Objetivo General

Clasificar y segmentar imágenes de microscopía electrónica de transmisión con escenarios de materiales nanoestructurados heterogéneos de silicato de itrio y óxido de silicio, usando inteligencia artificial con técnicas de aprendizaje profundo basadas en redes neuronales convolucionales integradas con una metodología para segmentación de instancias.

### 1.8.2 Objetivos Específicos

1. Crear un banco de datos de imágenes caracterizadas por microscopía TEM de silicato de itrio y óxido de silicio.
2. Diseñar una metodología que permita discriminar objetos superpuestos y traslapados.
3. Generar un banco de datos de imágenes de *ground truth* para las clases de silicato, silica, y coating.
4. Usar el banco de imágenes TEM y los *ground truth's* para entrenar la arquitectura Mask R-CNN.
5. Verificar el desempeño de la arquitectura computacional y la metodología con conjuntos de imágenes diferentes.

## Capítulo 2

# Metodología

### 2.1 Introducción

En este capítulo se aborda la preparación de las imágenes y las máscaras ”*ground truth*” con la metodología propuesta para resolver el problema de aglomeración y/o superposición de nanopartículas y materiales; y el preprocesado digital que se hizo a las imágenes para usarlas en la red neuronal. Se explican los tipos de aumentación de datos que se implementaron, las métricas de error y la arquitectura de la red neuronal.

### 2.2 Imágenes de Microscopía Electrónica de Transmisión

Las imágenes adquiridas por medio de la microscopía TEM (Cervantes *et al.* 2013; Flores *et al.* 2017) fueron inspeccionadas y revisadas visualmente (Figura 2-1) para verificar la calidad, claridad del escenario, resolución, y nitidez, seleccionando las mejores imágenes para el entrenamiento y para la preparación de los *ground truth*'s. De todo el conjunto de imágenes que se revisaron, se seleccionaron 80 muestras utilizadas para realizar la aumentación de datos, los *ground truth* y la entrada para la arquitectura Mask R-CNN. Derivado de la metodología que se desarrolló, con estas imágenes se generó un nuevo banco de imágenes que permitió tener una colección de 685 imágenes TEM (ver Tabla 2.1), con su par *ground truth* para las etapas de entrenamiento. En la fase de pruebas se usó un conjunto de 33 imágenes, el cual no fue aprendido por la red neuronal. Las imágenes TEM tenían una resolución original de  $2048 \times 2048$

Tabla 2.1: Distribución en la base datos de las tres clases de imágenes.

Clases	Número de	Porcentaje
Imágenes		
silicate	176	32%
silica	264	48%
coating	105	20%

pixeles, las cuales se cambiaron a  $256 \times 256$  pixeles. El reescalamiento de las imágenes se realizó con la biblioteca de OpenCV, usando el método de splines.

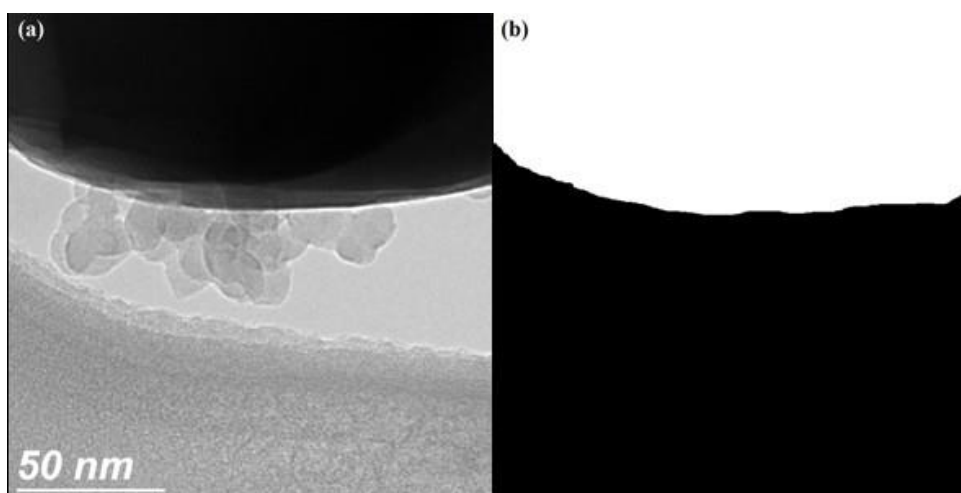


Figura 2-1: Ejemplo de una imagen TEM seleccionada para el entrenamiento. (a) La imagen muestra compuestos nanoestructurados con material denso de silicato de itrio (área más oscura). (b) Ground truth correspondiente al silicato de itrio.

### 2.2.0.1 Materiales luminiscentes

La creación de luz artificial ha sido un tema de investigación relevante durante las últimas décadas. Esto se debe principalmente a que estos nuevos dispositivos pueden lograr una reducción de hasta un 50% el consumo mundial de electricidad en iluminación, asimismo, una reducción de emisiones a la atmósfera, una mejor iluminación, mayor intensidad y lámparas más compactas (Sandia National Laboratories, 2006).

Las imágenes usadas en este trabajo fueron caracterizadas por microscopía TEM, y se generaron a partir de la síntesis de ortosilicato de itrio ( $Y_2 Si O_7$ ) dopado con iones de cerio ( $Ce^{3+}$ ) que emite en el azul, y iones de terbio ( $Tb^{3+}$ ) que tiene una emisión en el verde, al

Tabla 2.2: Se muestran las tres clases que se identifican en las imágenes por medio de metadatos en formato JSON.

Compuesto/Nanoestructura	Clase
Silicato de Itrio	Silicate
Oxido de Silicio	Silica
Recubrimiento del material	Coating

incorporar simultáneamente ambas tierras raras en el silicato de itrio se genera un material que emite luz blanca. Estos materiales se recubren con depósitos de sílice en forma de películas delgadas antirreflejantes, el cual mejora las propiedades luminiscentes del material sintetizado (Cervantes 2014).

### 2.3 Creación de *Ground Truth's*

El término “*ground truth*” es un resultado ideal esperado, sin embargo es la base de comparación usando la intersección sobre la unión para determinar que tanto un objeto se acerca a la realidad verdadera. Todas las imágenes fueron creadas y editadas en formato JPEG. El software GIMP 2.10.8 (GNU Image Manipulation Program) se usó para generar las imágenes de *ground truth*.

Para la creación de estas imágenes se separaron en tres clases diferentes: *silicate*, *silica* y *coating*, como se muestra en la Figura 2-2. Los archivos de texto de metadatos en formato JSON describen la clase a la que pertenece cada imagen (Tabla 2.2).

Los métodos tradicionales de segmentación no permiten el traslape de objetos, o discriminar objetos que están atrás de otros que poseen un cierto nivel de transparencia. En este sentido, la metodología que se implementó para lograr identificar objetos aglomerados, superpuestos o traslapados en principio fue discriminando visualmente todas las posibles nanopartículas de óxido de silicio que se encuentran en alguna de estas situaciones: aglomerada, superpuesta y/o traslapada. En el caso de la Figura 2-2a se puede observar una gran cantidad de nanopartículas, las cuales se encuentran en alguna de las situaciones que se tienen en consideración.

Cada imagen TEM que contiene nanopartículas de óxido de silicio (silica) fue separada en varias imágenes con un *ground truth* diferente, y que no traslapa con ninguna nanopartícula existente en otra imagen *ground truth* de la misma imagen. En este sentido, se pudieron extraer tantas imágenes como nuestra vista lo permitió al discriminar en un plano bidimensional

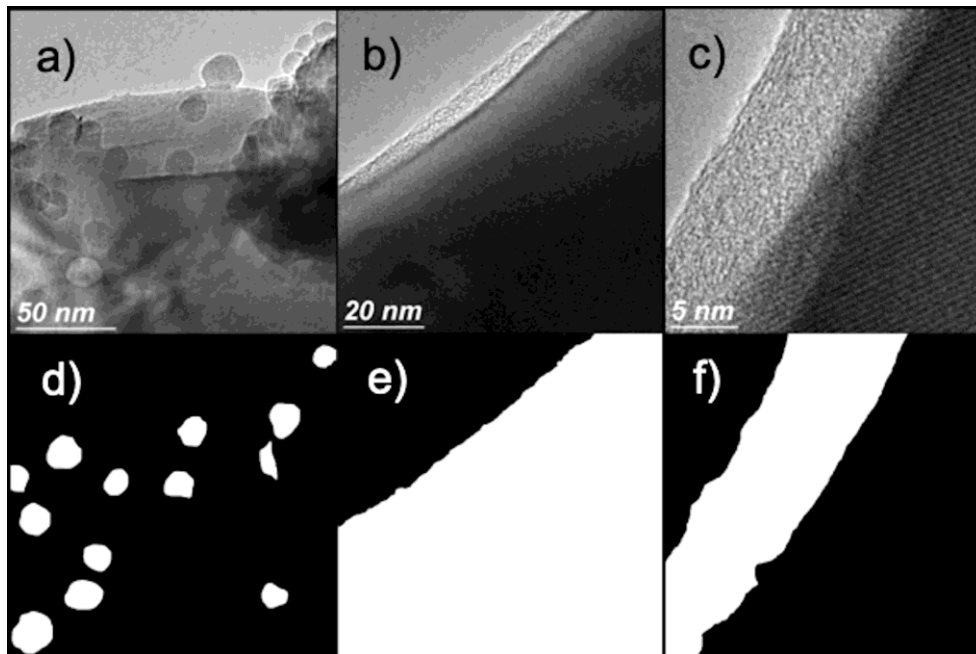


Figura 2-2: Se presentan las tres clases distintas usadas para el entrenamiento. Los recuadros a,b,c son las imágenes TEM, y los recuadros d,e,f son los *ground truth* creado a partir de las imágenes TEM. a) clase silica, b) clase silicate, y c) clase coating.

la profundidad de la superposición o niveles de nanopartículas en el escenario. Si se observan tres nanopartículas superpuestas o traslapadas entre las nanopartículas y/o el material heterogéneo de silicato en una imagen TEM, entonces se obtienen tres imágenes de ground truth, y su par TEM es la misma imagen original (Figura 2-3 y 2-2a,d) para los tres ground truth obtenidos. De esta forma, cada ground truth mantendrá nanopartículas diferentes para ser identificadas en su nivel de profundidad.

### 2.3.1 Métricas de evaluación

Las métricas para la precisión que se usaron tienen una función muy importante en la evaluación de los resultados de los modelos de segmentación. Los resultados obtenidos se obtuvieron a partir de la regla de la intersección sobre la unión calculada con la métrica del coeficiente de Dice, así como la sensibilidad y la especificidad.

$$\text{Coeficiente de Dice} = \frac{2TP}{2TP + FP + FN}, \quad (2.1)$$

$$\text{Sensibilidad (TVP)} = \frac{TP}{TP + FN}, \quad (2.2)$$

$$\text{Especificidad (TVN)} = \frac{TN}{TN + FP} \quad (2.3)$$

El coeficiente de Dice es usado como un índice el cual estima o mide el traslape o superposición entre el ground truth y la predicción. La sensibilidad, conocida como la Tasa de Verdaderos Positivos (TVP) es la proporción de casos positivos que fueron correctamente identificados. La especificidad, conocida como la Tasa de Verdaderos Negativos (TVN) es la proporción de casos negativos que el algoritmo identificó correctamente.

## 2.4 Ecuación de imágenes

Los histogramas muestran la distribución de la intensidad de los píxeles en una imagen y la ecualización de un histograma mejora el contraste de la imagen, logrando tener una distribución uniforme de los niveles de gris. Es una técnica ampliamente usada en el procesamiento digital de imágenes para mejorar los contrastes, dando como resultado un uso muy amplio de estos

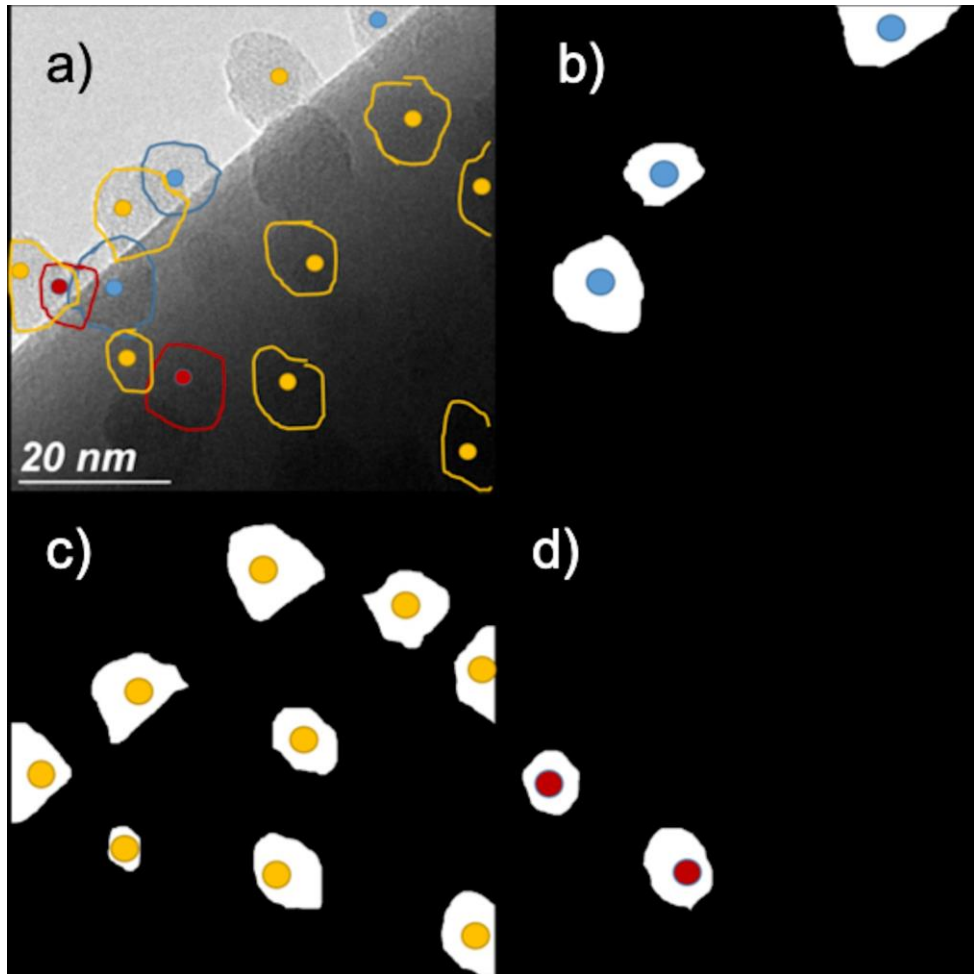


Figura 2-3: El recuadro (a) presenta la imagen TEM, sobre la cual observamos a simple vista nanopartículas superpuestas y traslapadas. Se muestran hasta tres niveles de profundidad en nanopartículas de óxido de silicio en los recuadros b,c,d, lo cual genera tres imágenes de ground truth.

métodos en distintas áreas de la ciencia como la industria aeroespacial, robótica, diseño e impresión de circuitos, biología, imágenes médicas, oceanografía satelital, industria, entre otros (Vidyasaraswathi and Hanumantharaju 2015; Vidhya and Ramesh 2017; Zhang *et al.* 2018).

La técnica de Contrast-Limited Adaptive Histogram Equalization (CLAHE) es un método de mejoramiento de contraste en imágenes médicas como rayos x (Rx) (Vidhya and Ramesh 2017). CLAHE se diferencia del método tradicional de ecualización de histograma adaptativo (AHE por sus siglas en inglés) por su limitación de contraste (CL por sus siglas en inglés). En CLAHE, el procedimiento de limitación de contraste se aplica para cada vecindad de la que se deriva una función de transformación. CLAHE previene la sobreamplificación del ruido de una imagen, esto permite una ecualización adaptativa del histograma, limitando el realce de contraste de AHE (Vidhya and Ramesh 2017). La amplificación del contraste en las proximidades de un pixel viene dada por la pendiente de la función de transformación, siendo proporcional a la pendiente de la función de distribución acumulativa de la vecindad (CDF) y, por lo tanto, al valor del histograma en el pixel.

CLAHE limita la amplificación del histograma a un valor umbral de corte antes de calcular la CDF (Vidhya and Ramesh 2017), limitando la pendiente de la CDF y, de la función de transformación. El valor al que se recorta el histograma, el llamado límite de recorte, depende de la normalización del histograma y, por tanto, del tamaño de la región de vecindad. Sin embargo, el límite de corte en CLAHE es invariante al contenido de la imagen

$$corte = \frac{N_b}{L} + \alpha \left( N_b - \frac{N_b}{L} \right),$$

donde  $N_b$  es el número de pixeles en cada sub-bloque,  $L$  es el factor de corte en el dominio de  $[0, 1]$  y  $G$  son los niveles de gris.

La parte del histograma que excede el límite de corte no se descarta, sino que es redistribuido por igual entre todos los contenedores de histogramas (Figura 2-4). Esta redistribución empujará algunos contenedores por encima del límite de corte nuevamente  $H_{i,j}$  (región en verde en la Figura 2-4), lo que dará como resultado un límite de recorte efectivo que es mayor que el límite prescrito y cuyo valor exacto depende de la imagen. Si esto no es deseable, el procedimiento de redistribución  $\tilde{H}_{i,j}$  se puede repetir de forma recursiva hasta que el exceso sea insignificante.

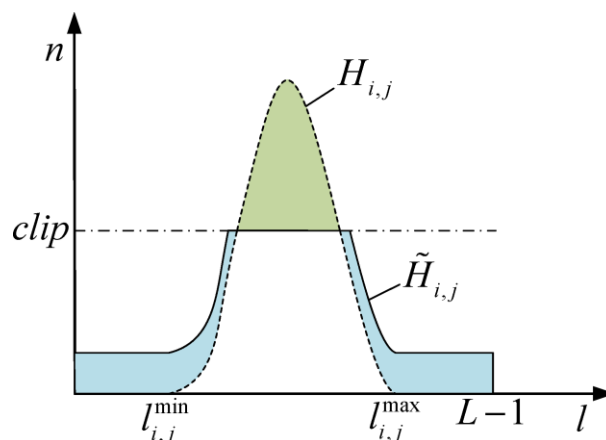


Figura 2-4: Ajuste del histograma redistribuido  $H_{i,j}$  (Liu *et al.* 2019)

Tabla 2.3: Se presentan los métodos de aumentación aplicados a las imágenes aleatoriamente, así como el rango que aplicó el método.

Método	Aumentación
Crop	0-10%
Scale	80-120%
Translate	-20 a 20
Rotate	-45 a 45
Shear	-16 a 16
Cval	0 a 255
Superpixels	20 a 200
Dropout	0.01 a 0.1

## 2.5 Aumentación de datos

Las aplicaciones de redes neuronales convolucionales tienen una demanda alta de datos para el entrenamiento. En visión computacional este tipo de aprendizaje tiene un desempeño bastante sobresaliente. Sin embargo, en muchos de los casos como biomedicina o nanotecnología no hay acceso a grandes volúmenes de datos. Para estos casos donde la falta de datos es una prioridad, se recurre a la aumentación de datos (Shorten and Khoshgoftaar 2019).

Los esquemas de aumentación de datos aplicados en tiempo de ejecución se presentan en la Tabla 2.3. Esto permitió tener un incremento en la base de datos y se redujo el sobreentrenamiento.

Tabla 2.4: Se presentan los hiperparámetros usados durante el entrenamiento de 100 épocas

Hiperparámetros	Valor
Batch Size	4
Steps Per Epoch	500/1000
Images Per GPU	4
Image Size	256 × 256
Learning Rate	0.0001
Detection NMS Threshold	0.9
RPN NMS Threshold	0.5
Mean Pixel	(128.0, 128.0, 128.0)
BackBone	ResNet 101
Gradient Clip Norm	10.0
Learning Momentum	0.9
Weight Decay	0.00005
Detection Min Confidence	0.5
ROI Positive Ratio	0.66

## 2.6 Hiperparámetros

Los hiperparámetros se fijaron y probaron con diferentes tipos de valores evaluando el desempeño de la arquitectura. Los valores presentados en la Tabla 2.4 son los hiperparámetros que presentaron el mejor desempeño de la arquitectura, resolviendo los objetivos críticos del proyecto, el detectar nanopartículas aglomeradas y/o superpuestas.

Un “batch size” grande puede resultar en un mejor gradiente, sin embargo también en un alto consumo de memoria. El batch size adecuado es de 4 imágenes. El hiperparámetro de “*Steps per Epoch*” se define como la cantidad de imágenes dividido por el “batch size”.

$$\text{PasosPorÉpoca(StepsPerEpoch)} = \frac{\text{Train Samples}}{\text{Batch Size}}$$

Debido a que el banco de datos fue aumentado como se indicó en la sección anterior, se pudo tener el hiperparámetro de steps per epoch fijado en 500 y 1000 para el entrenamiento. La Tabla 2.5 presenta el tiempo de ejecución para cada época.

Tabla 2.5: Tiempo de ejecución de la arquitectura neuronal durante el entrenamiento.

Pasos por época	Tiempo por época
500	240 s
1000	400 s

## 2.7 Entrenamiento

La Figura 2-5 muestra los tres tipos principales de capas de neuronas en un modelo de aprendizaje profundo: capa de entrada, capa(s) ocultas, y capa de salida. Cada conexión entre neuronas esta asociada con una matriz de pesos para transferir el conocimiento a la siguiente capa. Al entrenar la red, cuando viajan las sinapsis hacia adelante, se le enseña a la red, y cuando corre hacia atrás (se propaga el error) la red aprende, y los pesos son actualizados con el nuevo conocimiento adquirido. El modelo entrenado, con las matrices de pesos actualizadas, puede predecir las clases para las que se entrenó sobre nuevos datos de imágenes no etiquetados.

La arquitectura neuronal supervisada se entrenó con los pesos iniciales del conjunto de datos de MS COCO<sup>1</sup>. COCO es una base de datos de imágenes de gran escala usada para detección y segmentación de objetos; lo integran aproximadamente unas 330 mil imágenes con más de 200 mil etiquetadas, 1.5 millones de instancias de objetos y 80 clases de objetos. Estas imágenes se entrenan con sus categorías y etiquetas, y producen transferencia de conocimiento con este conjunto de pesos entrenados con sus imágenes. Los pesos en una red neuronal convolucional se pueden fijar con valores arbitrarios, o bien con matrices de pesos de otros entrenamientos, o con transferencia de conocimiento que en este caso se usaron los pesos del conjunto de datos entrenados de COCO. Sin embargo el usar valores arbitrarios, puede llevar a un entrenamiento bastante largo en tiempo de GPU y CPU.

### 2.7.1 Cómputo con GPU

En este entrenamiento se usó una computadora de escritorio con un procesador Intel i9-9900 con 16 núcleos, 16 Gb RAM y una tarjeta NVIDIA RTX 2080-Ti con 4352 cores y 11 Gb de RAM. Con esta característica de la memoria de la tarjeta NVIDIA se calculó el batch size.

---

<sup>1</sup><http://www.cocodataset.org>

### 2.7.1.1 TensorFlow

Tensorflow es un framework de código abierto desarrollado y liberado por Google™ que es usado para implementar algoritmos inteligentes de Machine Learning y Deep Learning. En ambas ramas tiene una serie de poderosos algoritmos que permiten a la computadora aprender patrones a partir de datos para tomar decisiones lo mejor posible.

### 2.7.1.2 Keras

Keras es uno de los frameworks de *Deep Learning* más usados en la actualidad. Permite con mucha facilidad y rapidez correr nuevos experimentos e ideas. Está construido encima de TensorFlow el cual puede escalar las aplicaciones a grandes clusters de GPU's o hasta un sistema total de TPU's.

En este proyecto se usó Keras como backend, lo cual permitió hacer ejecuciones de pruebas en computadoras integradas con solo el procesador intel, y con equipos integrados con tarjetas NVIDIA GTX así como RTX sin modificar el código.

### 2.7.1.3 Lenguaje de programación

Se usó el lenguaje de programación Python versión 3.7, y bibliotecas para manipular imágenes, datos y visualización como: numpy, scipy, matplotlib, opencv, json, glob, skimage, imgaug.

### 2.7.1.4 Sistemas operativos

El código se ejecutó sobre dos plataformas de sistemas operativos:

- Laptop con Windows 10, Intel i7, NVIDIA GTX 1080, configuración usada para pruebas.
- Desktop con Ubuntu 18.04, intel i9, NVIDIA RTX 2080 Ti, configuración usada para corridas finales. Los cálculos de la Tabla 2.5 se obtuvieron de este equipo.

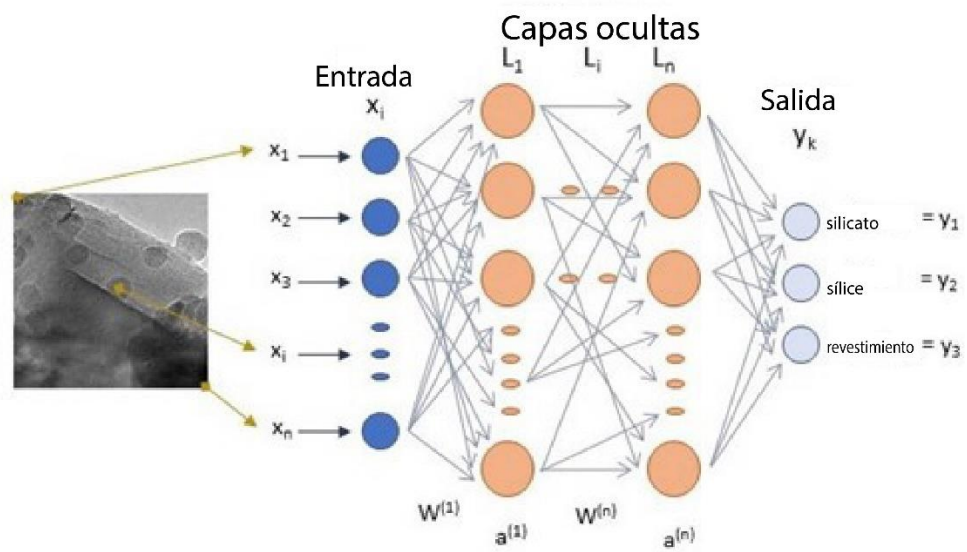


Figura 2-5: Diagrama esquemático de una red neuronal profunda (*Deep Learning*).

## Capítulo 3

# Resultados y Discusión

En este capítulo se presentan los resultados obtenidos con la metodología descrita anteriormente, mostrando las curvas de error durante el entrenamiento y las distribuciones de histogramas de los pesos entrenados. Además, se muestran las figuras donde las tres clases son identificadas por el sistema neuronal. Los hiperparámetros mostrados en la Tabla 2.4 son los usados en la configuración de la arquitectura, con los cuales se presentan los resultados obtenidos en este capítulo.

### 3.1 Entropía Cruzada Binaria

La entropía cruzada es una métrica de función de costo importante en aprendizaje profundo, porque ayuda a optimizar los modelos de clasificación, mientras menor sea la pérdida, mejor será el modelo. La entropía cruzada binaria (*binary cross-entropy*) define la métrica para la diferencia entre dos distribuciones de probabilidad para una variable aleatoria o un conjunto de eventos. Es usada en clasificación objetiva, y en segmentación al nivel de clasificación de pixel y en escenarios donde se tienen varios objetos para clasificar y segmentar. La entropía cruzada binaria para la  $k$  –ésima máscara correspondiente a la clase  $k$  del ground truth asociada a esta región, es definida por

$$L_{mask} = -\frac{1}{m^2} \sum_{1 \leq i, j \leq m} [y_{ij} \cdot \log \hat{y}_{ij}^k + (1 - y_{ij}) \cdot \log (1 - \hat{y}_{ij}^k)],$$

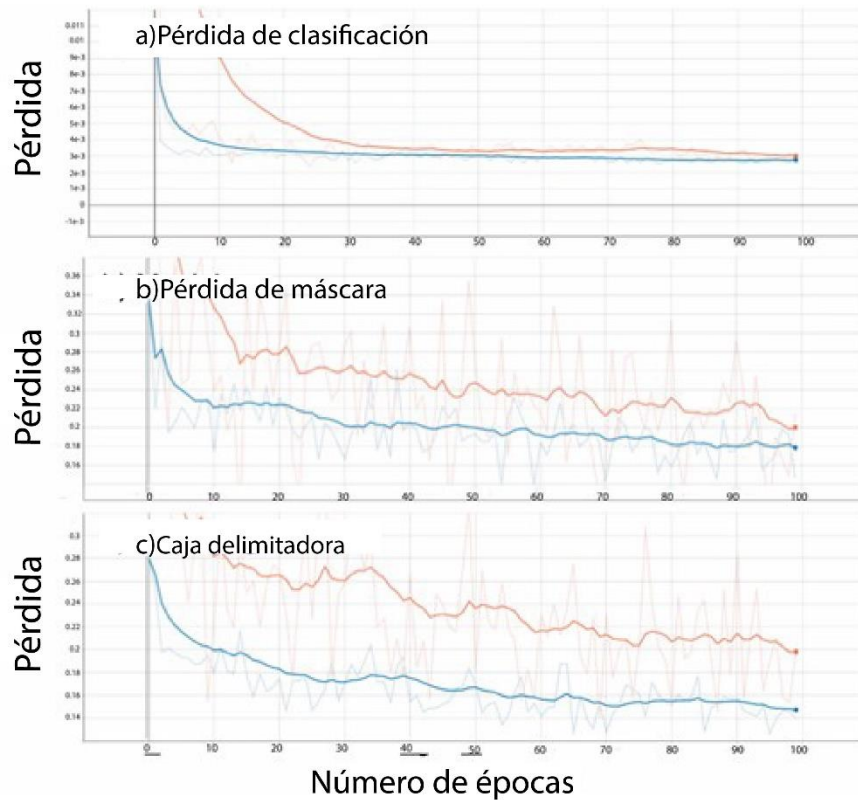


Figura 3-1: Funciones de pérdida de la arquitectura Mask R-CNN, se observa el desempeño de la red en: a) clasificación, b) segmentación, y c) caja delimitadora. La línea naranja muestra el entrenamiento realizado a 500 pasos por época y la línea azul a 1000 pasos por época.

donde  $y_{ij}$  son las etiquetas y  $\hat{y}_{ij}$  es la predicción para la  $k$  –ésima clase.

### 3.2 Entrenamiento

Los hiperpárametros de la Tabla 2.4 y los pasos por época indicados en la Tabla 2.5 se usaron para la corrida del entrenamiento de la arquitectura Mask R-CNN. La red supervisada se entrenó con los pesos inicializados con los del conjunto de datos de COCO. La función de pérdida se determina con la ecuación 3.1, las Figuras 3-1 y 3-2 muestran la convergencia a cero de la función de pérdida, significando esto que el error decrece y la red está aprendiendo de los datos; generando un mejor modelo de predicción profunda.

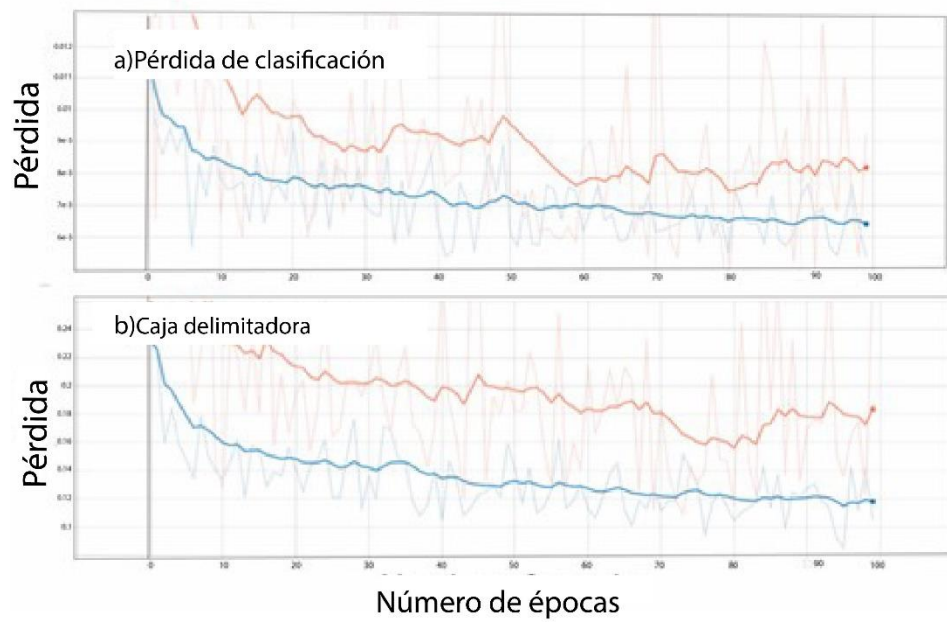


Figura 3-2: Funciones de pérdida de la arquitectura RPN, se observa el desempeño de la red en: a) clasificación, b) caja delimitadora. La línea naranja muestra el entrenamiento realizado a 500 pasos por época y la línea azul a 1000 pasos por época.

Tabla 3.1: Resultados de las funciones de pérdida de las redes Mask R-CNN y RPN después de 100 épocas con 500 y 1000 pasos por época.

Función de pérdida	500 pasos por época	1000 pasos por época	Descripción
mrcnn-class-loss	0.0030	0.0027	reconoce cada clase de objetos
mrcnn-bbox-loss	0.2017	0.1395	localiza los objetos
mrcnn-mask-loss	0.2153	0.1464	segmenta los objetos
rpn-class-loss	0.0092	0.0053	separa el fondo de los objetos
rpn-bbox-loss	0.2816	0.1055	localiza los objetos

El desempeño de la red neuronal en las 100 épocas (ver Tabla 3.1), muestra que a 1000 pasos por época la red genera un mejor modelo de predicción, reconociendo las clases de los objetos y localizándolos, además de lograr la capacidad de segmentar y separarlos del fondo de la imagen TEM. Las Figuras 3-1 y 3-2 muestran que el método del gradiente descendente logró encontrar los valores mínimos en cada época.

### 3.3 Matriz de pesos entrenados

Las Figuras 3-3 y 3-6 muestran la distribución en los histogramas de los pesos de los filtros entrenados en algunas de las capas de la arquitectura como ResNet101, Mask R-CNN y FPN. En todos los histogramas consistentemente se muestra una distribución gaussiana sobre todas las capas de red entrenables, mostrando que el entrenamiento de la red neuronal presenta isometría en las capas profundas, además que la incrustación de datos es estable durante todo el entrenamiento, lo que permite establecer una conexión entre la dimensión de las características y la complejidad de los datos (Giryes *et al.* 2016).

### 3.4 Detección, localización, clasificación y segmentación de múltiples compuestos

La Figura 3-7 presenta la inferencia de las tres clases del experimento sobre distintas imágenes TEM. De manera general la red predice con una precisión superior al 90% las tres clases. Los recuadros en la Figura 3-7a-c, muestran en color verde la nanopartículas de óxido de silicio, que son detectadas y clasificadas sobre distintos materiales como el silicato de itrio, o la rejilla

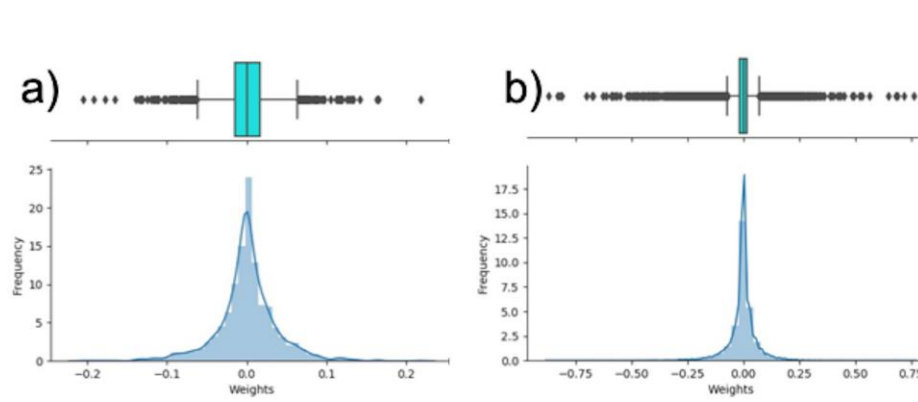


Figura 3-3: Distribución de matriz de pesos: a) Primera capa de convolución de la arquitectura ResNet 101, con un filtro de  $7 \times 7$ , con 3 canales y paso igual a 2, generando 64 mapas de características en la convolución. Los percentiles presentan un mínimo en  $-0.2053$ , máximo en  $+0.2185$  y una desviación estándar de  $+0.0385$ . El histograma representa la distribución de los pesos de los tensores en esta capa. b) Representa la capa de max pooling con un filtro de  $3 \times 3$  con paso igual a 2 para reducir la dimensionalidad de la capa beneficiando la invariancia en la traslación. El histograma representa la distribución de pesos de los tensores después del max pooling.

de vacío o sobre la película delgada (coating). Los recuadros c-e de la Figura 3-7, presentan la detección y clasificación de la clase silicate. La Figura 3-7c,f muestran la detección y clasificación de la clase coating.

Para algunos escenarios, donde hay nanopartículas vecinas el modelo se confunde y une como una sola nanopartícula los dos objetos. La clase "coating" debido al desbalance en la base de datos, es la que tiene menos representación para la arquitectura neuronal, lo que implica en resultados pobres de detección ya que no se tiene suficiente atención para la clase; esto crea un problema de validación y también en las imágenes de prueba. Sin embargo, el entrenamiento hace uso de aumentación de datos, lo cual resuelve en parte el problema de la falta de datos, pero el desbalance se manifiesta en la validación.

### 3.5 Detección y segmentación de compuestos traslapados y superpuestos

En la Figura 3-9 se puede observar que la red neuronal tiene la capacidad de detectar, localizar,

clasificar, y segmentar objetos traslapados y superpuestos de la clase silicon. Los recuadros a y b presentan similitud entre las dos segmentaciones, (a) representa la segmentación propuesta manualmente al ojo desnudo, y (b) es la segmentación inferida por la red neuronal.

Las técnicas de aprendizaje profundo son ampliamente usadas en visión computacional para detectar objetos, sobre todo en instancias que son difíciles de observar con rapidez, claridad y precisión. En el caso de la Figura 3-9 se tienen distintas instancias superpuestas que son complicadas a primera vista, sin embargo el modelo hace las inferencias correctas dada la complejidad de las estructuras.

Los objetos translucidos, opacos, o transparentes que intersectan con otros objetos traslapados u ocultos en su totalidad, tienen diferente saturación de píxeles, en blanco y negro o tonos de gris, y su histograma de distribución cambia de acuerdo a la profundidad del objeto que se localiza.

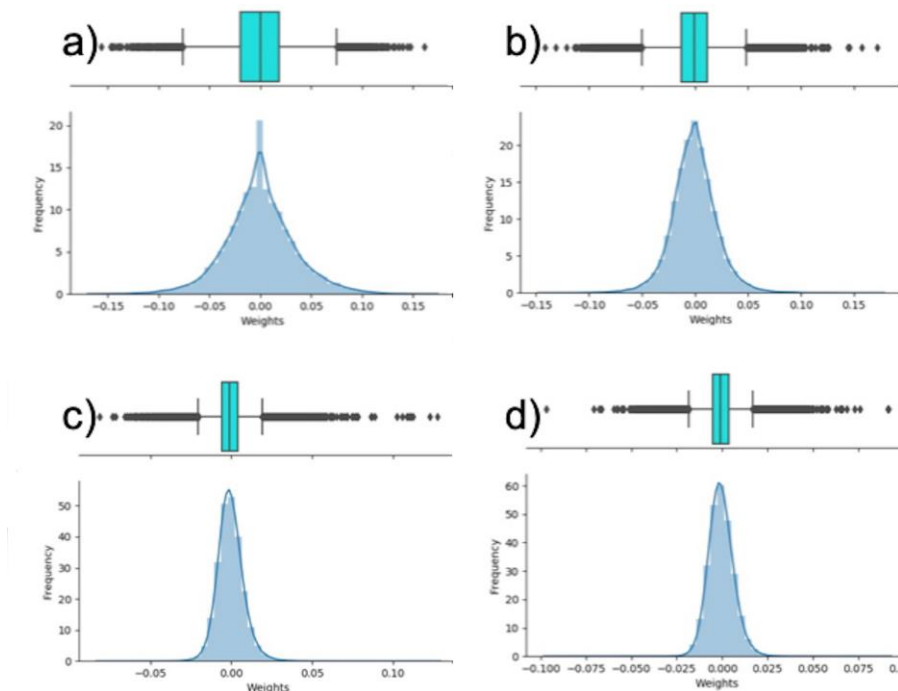


Figura 3-4: Distribución de matriz de pesos en ResNet 101: a) Convolución 2-c con filtro de  $1 \times 1$  y 64 canales, creando 256 mapas de características. b) Convolución 3-c con filtro de  $1 \times 1$  y 128 canales, creando 512 mapas de características. c) Convolución 4-c con filtro de  $1 \times 1$  y 256 canales, creando 1024 mapas de características. d) Convolución 5-c con filtro de  $1 \times 1$  y 512 canales, creando 2048 mapas de características.

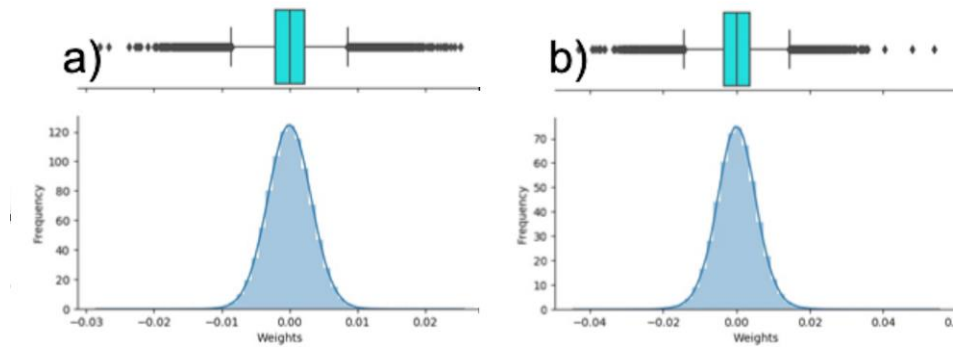


Figura 3-5: Distribución de pesos en Mask R-CNN para: a) predicción de la clase y b) predicción de la máscara. En ambos recuadros se presentan los cuartiles y los histogramas con la distribución de los pesos de los tensores.

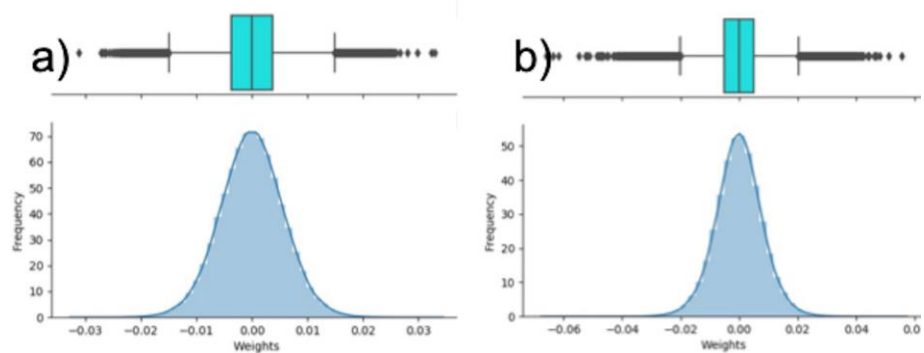


Figura 3-6: Distribución de la matriz de pesos de *Feature Pyramid Network* (FPN) para las capas: a) p5 y b) c5p5. Ambas capas crean una ruta descendente, la cual construye capas de alta resolución a partir de estructuras de alto nivel.

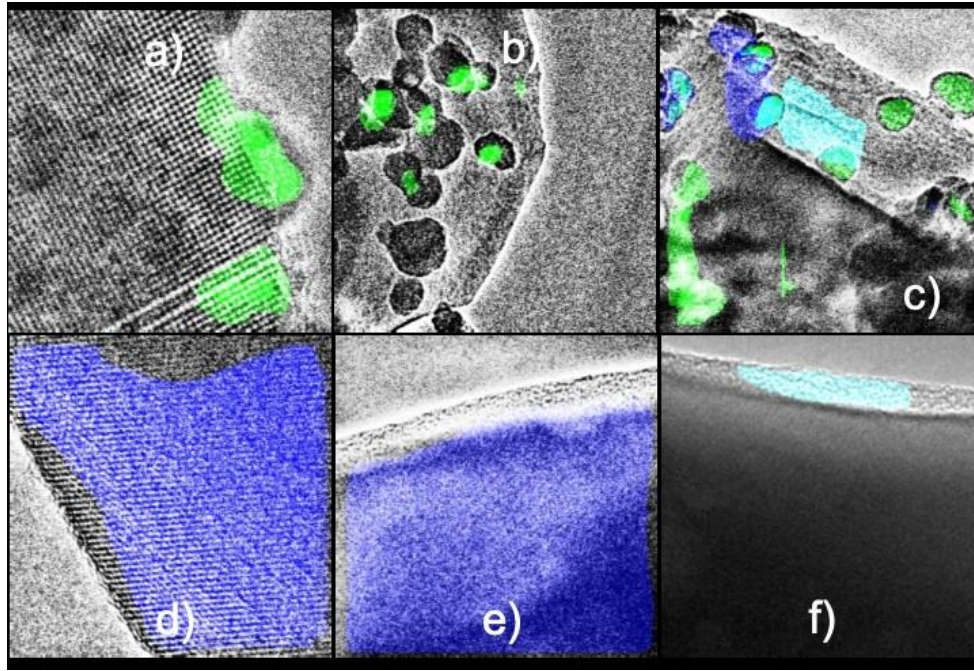


Figura 3-7: Segmentación de imágenes TEM para los tres tipos de clases: silicon (verde), silicate (azul) y coating (cyan). Los recuadros a-c muestran las predicciones de la red neuronal con segmentaciones de la clase silica (verde) con una precisión por encima del 92%. El recuadro c, particularmente muestra la clasificación y localización de las tres clases del experimento (silica, silicate, coating) con el modelo neuronal. Los recuadros d-e muestran la clase silicate inferida correctamente en imágenes a escalas en (d) a 10 nm y en (e) a 50 nm con una precisión por encima del 90%. El recuadro f presenta la segmentación en cyan correspondiente a la clase coating con una precisión por encima del 88%.

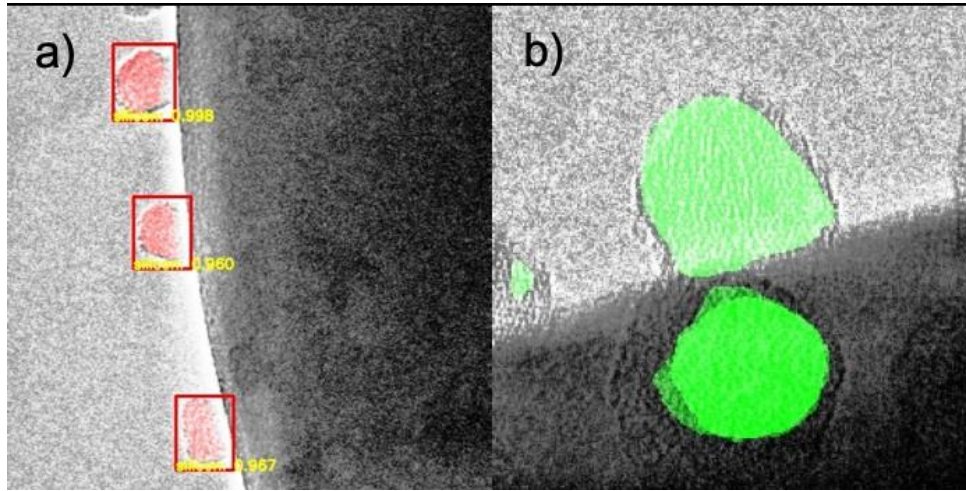


Figura 3-8: a) Nanopartículas de silicon detectadas, localizadas y clasificadas en la superficie del silicato de itrio; b) Nanopartículas de óxido de silicio detectadas, localizadas y clasificadas en la superficie del silicato de itrio y superpuestas al silicato de itrio, donde el modelo neuronal tiene la capacidad de detectar dos nanopartículas de silicon sobre la clase de silicate.

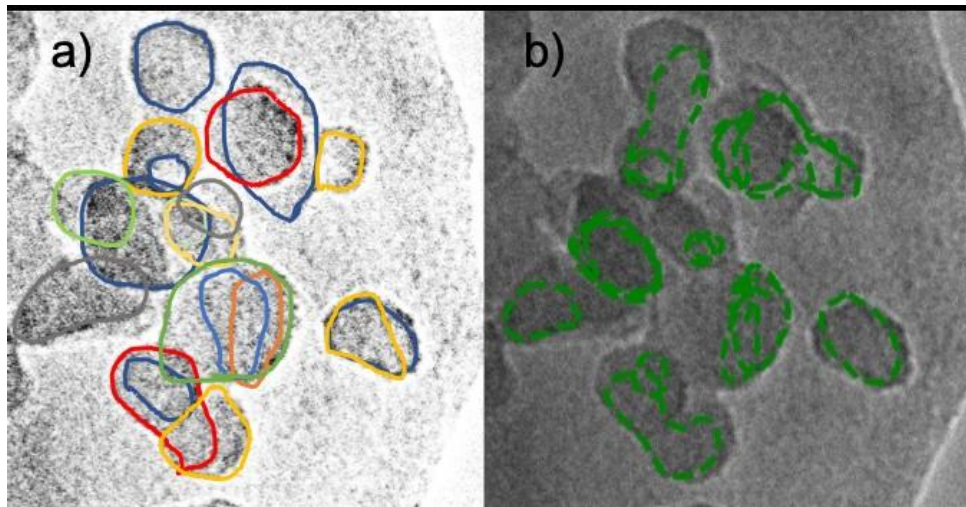


Figura 3-9: Esta figura muestra las nanopartículas de óxido de silicio (clase silicon) que están traslapadas o superpuestas. a) muestra las nanopartículas segmentadas a mano a simple vista y en b) se muestra la clase silicon como la red neuronal la detectó, localizó, clasificó y segmentó las nanopartículas.

## Capítulo 4

# Conclusiones y trabajo futuro

Este trabajo muestra que es posible detectar, localizar, clasificar y segmentar objetos, asignando un coeficiente de precisión a cada una de las clasificaciones; pero además con la capacidad de evaluar múltiples nanomateriales con estructura cristalográfica o productos de reacciones que se muestran traslapados y/o superpuestos en la caracterización por TEM. Estos traslapos, ocultamientos o superposiciones pueden llegar a tener varios niveles, y como nos muestra la metodología implementada, tiene la capacidad de predecir esos niveles de traslapos. La propuesta de este trabajo aporta una metodología para analizar materiales de escala microscópica (células, bacterias, organelos, cromosomas, etc.) y nanoscópica (nanoestructuras, cristales, etc.) de forma automática.

Se probaron distintas metodologías para generar los *ground truth*, y después de tener un banco de imágenes preparado se pasó a la fase de entrenamiento, validación y prueba. La metodología que se presenta en este trabajo permitió realizar detecciones de las clases de materiales propuestas con precisión sin confundir las clases. Sin embargo, debido al desbalance de la base de datos de imágenes se presenta una tendencia hacia la clase "silica" donde favorece las clasificaciones debido a que es la clase con más imágenes que alimentan la red neuronal.

La contribución de esta metodología al separar las nanopartículas aglomeradas y/o traslapadas tomadas de estos escenarios, básicamente está en las imágenes de *ground truth*, y tiene la capacidad de detectar tantos niveles de esas estructuras, como tantos niveles de capas de nanoestructuras o células superpuestas se separen en las imágenes de *ground truth*. El entrenamiento con la arquitectura Mask R-CNN y la metodología pudieron predecir,

localizando y segmentando las tres clases.

Esta metodología puede ser extendida a una amplia solución de problemas similares de dispositivos electrónicos que generan imágenes con estructuras ocultas o degradadas pero que mantienen información a nivel de pixel en la imagen. Algunas de las aplicaciones pueden ser en detección de cáncer de pulmón a partir de tomografía computarizada (CT), o con imágenes de resonancia magnética, detección de compuestos químicos a partir de imágenes de arreglos atómicos, o el crecimiento del recubrimiento de nanotubos, esto en el área de la química y la nanotecnología, sin embargo, puede extenderse al área de genómica en imágenes de ADN ya que son estructuras flexibles que se traslapan. También, en el área de protección civil, en la búsqueda de cuerpos y objetos perdidos como barcos y aviones hundidos en el océano. Esto incluso, aplica en la detección y localización de especies marinas.

Finalmente, la microscopía electrónica se aplica comunmente en la estimación del tamaño de partículas y sus distribuciones. Para obtener distribuciones de tamaño fiables es necesario determinar el tamaño de un gran número de partículas y, por tanto, es deseable una técnica computacional. Las aportaciones de trabajos en la detección de nanopartículas como Fisker *et al.* 2000 está limitado en aglomerados; unas décadas después Zafari *et al.* 2019 propone un nuevo método el cual requiere quitar el fondo de las imágenes que contienen las nanopartículas, detectando regiones convexas ajusta una elipse a esos puntos para cuantificar las nanopartículas. En Wang *et al.* 2021 donde proponen un método no supervisado para clasificar morfológicamente nanopartículas metálicas, depende de la geometría y el histograma de distribución definido de estos objetos. Aportaciones en aplicaciones biológicas para detectar objetos superpuestos (Phoulady *et al.* 2016; Li *et al.* 2016; Saleh *et al.* 2019; Böhm *et al.* 2019; Kowal *et al.* 2020) proponen métodos desde detección de bordes, de contornos, redes neuronales, distribución de histogramas y arquitecturas convolucionales. Unos tienen limitaciones en la localización de objetos muy pequeños debido al método, o la capacidad de solo detectar o clasificar objetos limitados.

En conclusión, la metodología propuesta en este trabajo permite detectar y clasificar objetos de cualquier tamaño, de distintas clases, además de no estar limitado geoméricamente. Como trabajo futuro, para mejorar la metodología se propone balancear e incrementar el tamaño de la base de datos.

# Referencias

- Affonso, Carlos, Debiaso Rossi, Andre Luis, Antunes Vieira, Fabio Henrique, de Leon Ferreira de Carvalho, Andre Carlos Ponce. Deep learning for biological image classification, *Expert systems with applications*. Volume: 85 Pages: 114-122, 2017. doi: 10.1016/j.eswa.2017.05.039.
- Affonso, C, Renato Jose Sassi, Ricardo Marques Barreirosa. Biological image classification using rough-fuzzy artificial neural network, *Expert Systems With Applications* 42 (2015) 9482–9488, 2015.
- Albrecht, Tim, Slabaugh, Gregory, Alonso, Eduardo, Al-Arif, S. M. Masudur R. Deep learning for single-molecule science, *Nanotechnology*. Volume: 28 Issue: 42, Article Number: 423001, 2017. doi: 10.1088/1361-6528/aa8334.
- A MongoDB white paper. Deep Learning and the Artificial Intelligence Revolution. April 2017.
- Antonio Hernandez Grande. Nanotecnología y nanopartículas magnéticas: La física actual en lucha contra la enfermedad, *Rev. R. Acad. Cienc. Exact. Fís. Nat. (Esp)* Vol. 101, N.2, pp 321-327, VIII Programa de Promoción de la Cultura Científica y Tecnológica, 2007.
- Captura de imagen en una camara digital. Tomado de Internet (julio 23, 2020): [https://www.fayerwayer.com / 2011 / 05 / captura-de-imagen-en-una-camara-digital-chwonders](https://www.fayerwayer.com/2011/05/captura-de-imagen-en-una-camara-digital-chwonders).
- Dormann J. & D. Fiorani, eds. *Magnetic Properties of Fine Particles*. North Holland, Amsterdam, 1992.
- Finnegan, Alex, Song, Jun S., Maximum entropy methods for extracting the learned features of deep neural networks. *PLOS Computational Biology*. Volume: 13 Issue: 10. Article Number: e1005836, 2017. doi: 10.1371/journal.pcbi.1005836.

- R. Fisker, R., J.M. Carstensen, M.F. Hansen, F. Bødker and S. Mørup. Estimation of nanoparticle size distributions by image analysis, *Journal of Nanoparticle Research* 2: 267–277, 2000.
- Fukushima, K. “Neocognitron: A Hierarchical Neural Network Capable of Visual Pattern Recognition”, *NeuralNetworks*, Vol. 1, pp. 119-130, 1988.
- Gadepally V. *AI Enabling Technologies: A Survey*, MIT, vol. 7014, 2019.
- Gal Yarin, Riashat Islam, Zoubin Ghahramani. Deep Bayesian Active Learning with Image Data, arXiv:1703.02910v1 [cs.LG], 2017.
- Ganesh R. Vidhya and H. Ramesh. Effectiveness of Contrast Limited Adaptive Histogram Equalization Technique on Multispectral Satellite Imagery. In *Proceedings of the International Conference on Video and Image Processing (ICVIP 2017)*. Association for Computing Machinery, New York, NY, USA, 234–239, 2017. doi: <https://doi.org/10.1145/3177404.3177409>
- Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghahfouri, Jeroen A.W.M. van der Laak, Bram van Ginneken, Clara I. Sanchez. A Survey on Deep Learning in Medical Image Analysis, *Medical Image Analysis*, 2017. doi: 10.1016/j.media.2017.07.005.
- Giryès, R., G. Sapiro, and A. M. Bronstein. Deep Neural Networks with Random Gaussian Weights: A Universal Classification Strategy?, *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3444–3457, 2016. doi: 10.1109/TSP.2016.2546221.
- Goodfellow, I., Bengio, Y., and Courville, A. *Deep Learning*. MIT Press, 2016. <http://www.deeplearningbook.org>.
- Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Bengio, Y. Generative adversarial nets. *Advances in neural information processing systems* pp. 2672–2680, 2014.
- Gu, Jiuxiang, Zhenhua Wangb, Jason Kuenb, Lianyang Mab, Amir Shahroudyb, Bing Shuaib, Ting Liub, Xingxing Wangb, Li Wangb, Gang Wangb, Jianfei Caic, Tsuhan Chenc. Recent Advances in Convolutional Neural Networks, arXiv:1512.07108v6 [cs.CV], 2017.
- Hadjipanayis, G. and R. Siegel, eds. *Nanophase Materials. Synthesis – Properties – Applica-*

- tions. Kluwer Academic Publishers, Dordrecht, 1994.
- Hebb, D. O., and Hebb, D. The organization of behavior. Vol. 65. Wiley New York, 1949.
- Hochreiter, S., and Schmidhuber, J. Long short-term memory. *Neural computation*, 9(8), 1735–1780, 1997.
- Horwath J.P., Dmitri N. Zakharov, Rémi Mégret and Eric A. Stach. Understanding important features of deep learning models for segmentation of high-resolution transmission electron microscopy images, *Computational Materials* 6:108, 2020. <https://doi.org/10.1038/s41524-020-00363-x>
- Inkson, B.J. 2 - Scanning electron microscopy (SEM) and transmission electron microscopy (TEM) for materials characterization, Editor(s): Gerhard Hübschen, Iris Altpeter, Ralf Tschuncky, Hans-Georg Herrmann, *Materials Characterization Using Nondestructive Evaluation (NDE) Methods*, Woodhead Publishing, 2016, Pages 17-43, ISBN 9780081000403, <https://doi.org/10.1016/B978-0-08-100040-3.00002-X>
- Kowal, M., M. Zejmo, M. Skobel, J. Korbicz, R. Monczak. Cell Nuclei Segmentation in Cytological Images Using Convolutional Neural Network and Seeded Watershed Algorithm. *Journal of Digital Imaging* 33:231–242, 2020.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324, 1998.
- LeCun, Y., Yoshua Bengio and Geoffrey Hinton. Deep Learning. *Nature* 521, 2015. doi:10.1038/nature14539
- Li Y., F. Rose, F. di Pietro, X. Morin and A. Genovesio. Detection and tracking of overlapping cell nuclei for large scale mitosis analyses. *BMC Bioinformatics* 17:183, 2016. doi:10.1186/s12859-016-1030-9
- Liu, C., Sui, X., Kuang, X., Liu, Y., Gu, G., Chen, Q. Adaptive Contrast Enhancement for Infrared Images Based on the Neighborhood Conditional Histogram. *Remote Sens.* 11, 1381, 2019. <https://doi.org/10.3390/rs11111381>.
- Madsen J., Pei Liu, Jens Kling, Jakob Birkedal Wagner, Thomas Willum Hansen, Ole Winther, Jakob Schiøtz. A deep learning approach to identify local structures in atomic

- resolution transmission electron microscopy images, arXiv:1802.03008v2, 2018.
- Mills, Kyle, Spanner, Michael, Tamblyn, Isaac. Deep learning and the Schrodinger equation, *Physical Review*. Volume: 96 Issue: 4. Article Number: 042113, 2017. doi: 10.1103/PhysRevA.96.042113.
- Natal, Rodrigo de Andrade, Pelegati, Vitor Bianchin, Bondarik, Caroline, Mendonca, Guilherme Rossi, Derchain, Sophie Françoise, Lima, Carmen Passos, Cesar, Carlos Lenz, Sarian, Lus Otavio, Vassallo, Jose. Increased metabolic activity detected by FLIM in human breast cancer cells with desmoplastic reaction: a pilot study. Edited by :Beaurepaire, E; So, PTC; Pavone, F; Hillman, EM. *Advanced Microscopy techniques IV; and Neutrophotonics II*. Book Series: *Proceedings of SPIE*. Volume: 9536. Article Number: 95360L, 2015. doi: 10.1117/12.2183442.
- Neubauer, C., Evaluation of convolutional neural networks for visual recognition, *IEEE Transactions on neural networks*, Vol. 9, No. 4, 1998.
- Ramcharan, Amanda, Baranowski, Kelsea, McCloskey, Peter, Ahmed, Babuali, Legg, James, Hughes, David P. Deep Learning for Image-Based Cassava Disease Detection, *Frontiers in plant science*, Volume: 8. Article Number: 1852, 2017. doi: 10.3389/fpls.2017.01852.
- Ribeiro E. and Shah M. Computer vision for nanoscale imaging, *Machine Vision and Applications* 17:147–162, 2006. doi: 10.1007/s00138-006-0021-7.
- H.M. Saleh, N.H. Saada, N.A.M. Isa. Overlapping Chromosome Segmentation using U-Net: Convolutional Networks with Test Time Augmentation. *Procedia Computer Science* 159 524–533, 2019.
- Samaniego, R. *Aplicaciones de la microscopia confocal*, Unidad de Microscopia Confocal. Hospital General Universitario Gregorio Marañón. Madrid. España, 2009.
- Shorten c., and T. M. Khoshgoftaar. A survey on Image Data Augmentation for Deep Learning, *J. Big Data*, 2019, doi: 10.1186/s40537-019-0197-0.
- Turing, A. On Computable Numbers, With An Application To The Entscheidungsproblem, In *Proceedings of the London Mathematical Society, Series 2*, Vol. 42, pp. 230-265, 1936.
- Vidyasaraswathi H.N. and M. C. Hanumantharaju. Review of Various Histogram Based Medical

Image Enhancement Techniques. In Proceedings of the 2015 International Conference on Advanced Research in Computer Science Engineering & Technology (ICARCSET 2015). Association for Computing Machinery, New York, NY, USA, Article 48, 1–6, 2015. DOI:<https://doi.org/10.1145/2743065.2743113>

Wang, X., J. Li, H.D. Ha, J.C. Dahl, J.C. Ondry, I. Moreno-Hernandez, T. Head-Gordon, and A.P. Alivisatos. AutoDetect-mNP: An Unsupervised Machine Learning Algorithm for Automated Analysis of Transmission Electron Microscope Images of Metal Nanoparticles, *JACS Au*, 1, 316-327, 2021.

Watkins, C. J., and Dayan, P. Q-learning. *Machine learning*, 8(3-4), 279–292, 1992.

Werbos, Paul. Neural Networks that actually work in diagnostics, prediction and control: Common misconceptions vs real-world success. Tutorial IEEE Computational Intelligence Society. 2005.

Zhang L., Huan Jin, and Dongmei Wu. Histogram Equalization Based on Custom Region of Interest. In Proceedings of the 3rd International Conference on Robotics, Control and Automation (ICRCA '18). Association for Computing Machinery, New York, NY, USA, 185–189, 2018 doi: <https://doi.org/10.1145/3265639.3265678>

Zhao, Dekuang, Tahaney, William M., Mazumdar, Abhijit, Savage, Michelle I., Brown, Powel H. Molecularly targeted therapies for p53-mutant cancers. *Cellular and molecular life sciences*. Volume: 74 Issue: 22 Pages: 4171-4187, 2017. doi: 10.1007/s00018-017-2575-0.

Zhen, Xin, Chen, Jiawei, Zhong, Zichun, Hrycushko, Brian, Zhou, Linghong, Jiang, Steve, Albuquerque, Kevin, Gu, Xuejun. Deep convolutional neural network with transfer learning for rectum toxicity prediction in cervical cancer radiotherapy: a feasibility study, *Physics in medicine and biology*, Volume: 62 Issue: 21 Pages: 8246-8263, 2017. doi: 10.1088/1361-6560/aa8d09.