



UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO
PROGRAMA DE MAESTRÍA Y DOCTORADO EN INGENIERÍA
INGENIERÍA ELÉCTRICA – PROCESAMIENTO DIGITAL DE SEÑALES

DETECCIÓN, LOCALIZACIÓN Y SEGMENTACIÓN DE CEREBELO EN IMÁGENES DE
ULTRASONIDO FETAL

TESIS
QUE PARA OPTAR POR EL GRADO DE:
MAESTRO EN INGENIERÍA

PRESENTA:
RODRIGO RAMOS DÍAZ

TUTORA PRINCIPAL
JIMENA OLVERES MONTIEL, FACULTAD DE INGENIERÍA

Ciudad de México, diciembre 2020



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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

JURADO ASIGNADO:

Presidente:	Dr. Gibran Fuentes Pineda
Secretario:	Dr. Boris Escalante Ramírez
1 ^{er} Vocal:	Dra. Jimena Olveres Montiel
2 ^{do} Vocal:	Dr. Miguel A. Padilla Castañeda
3 ^{er} Vocal:	Dr. Fernando Arámbula Cosío

TUTOR DE TESIS:

JIMENA OLVERES MONTIEL

FIRMA

Agradecimientos

**Se agradece al CONACYT por mi beca de maestria,
al Programa de Apoyo a Proyectos de Investigación e Innovación Tecnológica UNAM-
PAPIIT IA103119 e IV100420 de la UNAM,
así como a la Secretaria de Educación, Ciencia, Tecnología e Innovación SECTEI
202/2019, por el apoyo concedido a este trabajo.**

Abstract

The discovery of ultrasound in bats to avoid obstacles in their night flight and to hunt was an extraordinary event in its time. However, it was until the middle of the 20th century that this fact was confirmed and since then work and research have been carried out to apply this type of energy in medicine. It has been particularly useful and has become a very common practice in OB / GYN for monitoring pregnancy. The ultrasound images of the fetus and specifically of the fetal brain serve among other purposes to estimate gestational age.

On the other hand, the study of the brain, nervous system and behavior in humans and other animals has inspired an artificial intelligence engineering and a group of cognitive sciences that is concerned with investigating the mind and its processes. In recent times, 2000s, major advances have been made in machine learning with multi-layered neural networks or deep networks. Towards the end of the 2010s, research in deep learning applications for medicine, mainly medical imaging, has taken off.

The objective of this thesis was to implement the detection, localization, and segmentation of the cerebellum in fetal ultrasound images (transcerebellar plane). A review was made of the state of the art in deep learning applied to medical ultrasound. Two known architectures were tested and some of their hyperparameters were experimented with, such as the number of convolutional layers and the output size for a dense layer. Two aspects of machine learning were also experimented with, namely the initialization and optimization methods. These experiments allowed to improve the performance of the implemented tasks.

Resumen

El descubrimiento de los ultrasonidos en los murciélagos para evitar obstáculos en su vuelo nocturno y cazar fue un acontecimiento extraordinario en su tiempo. Sin embargo, fue hasta mediados del siglo XX que se confirmó este hecho y desde entonces se han realizado trabajos e investigaciones para aplicar este tipo de energía en la medicina. Ha resultado particularmente útil y se ha convertido en una práctica muy común en gineco-obstetricia para el seguimiento del embarazo. Las imágenes por ultrasonido de feto y en específico de cerebro fetal sirven entre otros propósitos para estimar la edad gestacional.

Por otra parte, el estudio del cerebro, el sistema nervioso y la conducta en humanos y otros animales ha inspirado una ingeniería de inteligencia artificial y un grupo de ciencias cognitivas que se ocupa de investigar la mente y sus procesos. En tiempos recientes, años 2000, se han hecho importantes avances en el aprendizaje automático con redes neuronales de muchas capas o redes profundas. Hacia finales de los años 2010 se ha visto un despegue de la investigación en aplicaciones de aprendizaje profundo para medicina, principalmente, imágenes médicas.

El objetivo de esta tesis fue implementar la detección, localización y segmentación de cerebelo en imágenes de ultrasonido fetal (plano transcerebelar). Se hizo una revisión del estado del arte en aprendizaje profundo aplicado a ultrasonido médico. Se probaron dos arquitecturas conocidas y se experimentó con algunos de sus hiperparámetros como el número de capas convolucionales y el tamaño de salida para una capa densa. También se experimentó con dos aspectos del aprendizaje automático que son los métodos de inicialización y de optimización. Dichos experimentos permitieron mejorar el desempeño de las tareas implementadas.

Contenido

Abstract	5
Resumen.....	6
Índice de figuras	9
1. Introducción	12
1.1 Justificación	12
1.2 Objetivos	12
1.3 Hipótesis.....	13
2. Ultrasonido médico y redes neuronales	15
2.1 Historia del Ultrasonido	15
2.2 Ultrasonido y cerebelo fetal.....	20
2.3 Inteligencia Artificial y Ciencias Cognitivas	22
2.4 Redes Neuronales Artificiales y Aprendizaje Profundo.....	25
3. Análisis automático de imágenes médicas.....	29
3.1 Detección.....	29
3.2 Clasificación.....	30
3.3 Segmentación.....	31
3.4 Análisis 3D	35
4. Discusión	38
4.1 Alcances.....	38
4.2 Antecedentes	38
4.3 Estado del arte	39
5. Metodología	42
5.1 Datos y métricas.....	42
5.2 Descenso por gradiente	43
5.3 Detección y localización	43
5.4 Segmentación semántica	46
6. Resultados	50
6.1 Detección y localización	50
6.2 Segmentación semántica	54

7. Conclusiones.....	59
Bibliografía	63

Índice de figuras

Escáner con brazo articulado [2].....	17
Escáner con bolsa de agua [2].....	17
Cefalometría de Campbell. El modo A (izquierda) permitía hacer mediciones precisas y el modo B (derecha) permitía visualizar [3].....	18
Vistas axiales de la cabeza fetal. (a) Plano transventricular; (b) plano transtalámico; (c) plano transcerebelar [5].....	21
Vista transversal de cabeza fetal (semana 26) demostrando encefalocele. La herniación de meninges y cerebelo son características del síndrome Arnold-Chiari III [7].....	22
Ciencias Cognitivas e Inteligencia Artificial	24
Segmentación de VI fetal con CNN dinámicas [77].....	32
Modelo apilado de FCN que produce un resultado intermedio con el objeto buscado y otros similares [58].....	33
Segmentación de hueso con aprendizaje profundo. Registro, emparejamiento y propagación de información preoperatoria a imagen por US en el quirófano [78].	33
La estimación de la forma se refina de forma gradual mediante RNN para completar bordes (BCRNN) [42].	34
Implantación de FCN en cascada mostrando dos niveles de autocontexto [79].....	35
Segmentaciones de endometrio con U-Net y U-Net más propagación de curvas, en rojo y amarillo respectivamente [80].....	36
Características de Hadamard y aprendizaje profundo para tareas de clasificación y localización [81].	36
Aplicación de FCN 3D y RNN para segmentación de volúmenes [82].....	37
Adquisición y procesamiento de señales ultrasónicas en un sistema CADx.....	38
Visión amplia de los fundamentos para el análisis de imágenes médicas mediante aprendizaje profundo.....	39
Palabras clave más relevantes en nuestro resumen del estado del arte.....	40
Ejemplo de máscara de segmentación con el cerebelo en blanco y su imagen de US correspondiente.....	42
Arquitectura YOLO con 24 capas convolucionales.....	44
Representación gráfica del tensor de salida [99].....	45

Primeras tres versiones de redes totalmente convolucionales para segmentación semántica [102].	47
Contornos a partir de las anotaciones.	49
Grupos de cuartos de contorno proximal (izquierda) y distal (derecha).	49
Evaluación de localización la izquierda (precisión promedio) y detección a la derecha (puntaje F1).....	50
Inferencia con modelo de 49 divisiones, y después de 500 lotes, a la izquierda. Anotación experta a la derecha.....	51
Inferencia con modelo de 81 divisiones, y después de 600 lotes, a la izquierda. Anotación experta a la derecha.....	51
Inferencia con modelo de 121 divisiones, y después de 600 lotes, a la izquierda. Anotación experta a la derecha.....	52
Inferencia con modelo de 169 divisiones, y después de 1000 lotes, a la izquierda. Anotación experta a la derecha.....	52
Ejemplos de inferencias con falsos positivos.	54
Evolución de falsos positivos para dos tasas de aprendizaje.....	53
Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con datos no centrados, VGG19, optimizador Adadelta, y “desde cero”. El conjunto de pesos fue entrenado 400 épocas y obtuvo una evaluación AP=0.861.....	55
Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con datos centrados, VGG16, optimizador Adadelta, y con transferencia de aprendizaje. El conjunto de pesos fue entrenado 900 épocas y obtuvo una evaluación AP=0.947	55
Imagen de entrada, segmentación automática, y anotación experta para el mejor modelo obtenido por inicialización de Glorot. AP=0.918	56
Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con cuartos distales y aprendizaje desde cero. El conjunto de pesos fue optimizado 200 épocas mediante Adadelta y obtuvo una evaluación AP=0.914.....	57
Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con datos distales. El conjunto de pesos fue entrenado 200 épocas desde cero y obtuvo una evaluación AP=0.896.....	58
Propuesta de modelo apilado YOLO-FCN para segmentación fina del cerebelo.....	61
Propuesta de sistema automatizado para la biometría de cerebelo.....	61
Propuesta de modelo apilado YOLO-ASM.	61

Propuesta de modelo apilado FCN-ASM..... 61

1. Introducción

1.1 Justificación

El análisis y procesamiento de imágenes médicas, como parte del Procesamiento Digital de Señales, es una tarea importante por su relación con la salud pública. La obtención de imágenes médicas por ultrasonido es una práctica clínica que se ha hecho muy común gracias a su seguridad, comodidad, bajo costo, y resultados inmediatos. Sin embargo, las imágenes obtenidas son de calidad variable, contienen ruido característico al método de obtención, y requieren de personal altamente entrenado. Entonces sus beneficios todavía no son suficientemente accesibles. El examen prenatal por ultrasonido se prefiere por encima de rayos X, resonancia magnética y tomografía que presentan mayor riesgo para el feto. El cerebelo coordina la información sensorial y el control muscular, regula los estados emocionales y se investiga su participación en procesos cognitivos y lingüísticos. Se recomiendan exploraciones de cerebro fetal con ultrasonido en el segundo y tercer trimestre de gestación para detectar oportunamente algún problema en el desarrollo del sistema nervioso central. Existen malformaciones raras pero graves que afectan al cerebelo y pueden verse mediante ecografía. La biometría del cerebelo, también por US, se utiliza para determinar la edad gestacional cuando otros indicadores están en duda. En este trabajo se implementa la detección, localización, y segmentación de cerebelo en imágenes de US fetal mediante aprendizaje profundo. Dichas tareas se pueden ver como los primeros pasos hacia la automatización de la biometría para estimar la edad gestacional y también para construir un sistema inteligente que asista el diagnóstico de malformaciones.

1.2 Objetivos

El objetivo general es implementar detección, localización y segmentación de cerebelo en imágenes de ultrasonido fetal. La detección dirá si hay cerebelo o no en la imagen. La localización indicará el lugar del cerebelo en la imagen. La segmentación definirá el área que ocupa el cerebelo en la imagen.

- Investigar sobre los antecedentes y el estado del arte en aprendizaje profundo aplicado a imágenes médicas de ultrasonido.
- Implementar detección y localización de cerebelo en ecografías de cerebro fetal mediante aprendizaje profundo.
- Implementar la segmentación del cerebelo mediante aprendizaje profundo.
- Investigar si es factible mejorar la segmentación mediante la previa localización.
- Investigar si es factible mejorar la segmentación mediante un análisis del objeto.
- Experimentar con los hiperparámetros de las redes profundas empleadas y aspectos del aprendizaje como la inicialización de los parámetros y el método de optimización.

1.3 Hipótesis

- A. Una arquitectura profunda de redes neuronales artificiales (DNN), configurada y ajustada para detección y localización de objetos cotidianos (personas, animales, transportes, etc) en imágenes naturales, podría reajustarse para detectar y localizar estructuras anatómicas en imágenes de ultrasonido.
- B. Una red neuronal podrá aproximar cualquier función siempre y cuando tenga el número suficiente de elementos. Entonces podemos esperar que un mayor número de elementos en la red permitan mejorar el desempeño en la tarea aprendida.
- C. El aprendizaje mediante descenso por gradiente estocástico con momento tiene dos parámetros principales: la tasa de aprendizaje y el factor de momento. La tasa de aprendizaje debe tener el efecto más importante.
- D. Una DNN, configurada y ajustada para segmentación semántica de objetos cotidianos (personas, animales, transportes, etc) en imágenes naturales, podría reajustarse para segmentar estructuras anatómicas en imágenes de ultrasonido.
- E. Dicha DNN, configurada y ajustada para la segmentación semántica de un objeto en una imagen presentará mejores resultados cuando el objeto ocupe una parte central y preponderante en la imagen.
- F. Es posible entrenar o ajustar una DNN para segmentación con algunos cientos de ejemplos y sin aplicar transferencia de aprendizaje (con pesos inicialmente puestos en valores aleatorios).
- G. La DNN para segmentación semántica presentará mejores resultados cuando la segmentación se haga por partes y se conozcan diferentes modelos estadísticos de dichas partes.

2. Ultrasonido médico y redes neuronales

2.1 Historia del Ultrasonido

Es notoria la importancia que se le ha dado en este trabajo a la historia del ultrasonido médico. La intención ha sido de no obviar todo el desarrollo que se ha hecho en las últimas 8 décadas para obtener imágenes médicas por US. Por otra parte, una revisión de los fundamentos permite ampliar el horizonte de estudio.

Descubrimiento de los ultrasonidos

A finales del siglo XVIII Lázaro Spallanzani se interesó en la visión nocturna de algunos animales [1]. Observó que la lechuza y otros seres nocturnos perdían sus habilidades en un cuarto totalmente oscuro. En cambio, los murciélagos parecían capaces de ver las paredes y otros obstáculos aún en total oscuridad. Spallanzani llegó a remover los ojos de algunos murciélagos para comprobar que podían orientarse sin usar sus ojos. Luego liberó a algunos de estos especímenes cegados y descubrió que conservaban su capacidad para cazar insectos. Charles Jurine supo de estos experimentos y por su parte descubrió que estos animalitos perdían su habilidad si les taponaba las orejas con cera. Entonces Spallanzani repitió el experimento y luego insertó pequeños tubos de bronce que permitían la entrada del sonido al canal auditivo. Finalmente tapó los tubos con cera. Así se llegó a la entonces increíble conclusión de que los murciélagos utilizan su sentido del oído para cazar y evitar obstáculos en la oscuridad.

En el año de 1880 los hermanos Pierre y Jaques Curie descubrieron el efecto piezoeléctrico en cristales asimétricos, como el cuarzo y la sal Rochelle, que permiten crear transductores de vibraciones mecánicas a eléctricas (y viceversa) [2]. Estos transductores permitieron el desarrollo de dispositivos ultrasónicos y de la electrónica de precisión. Paul Langevin y Constantin Chilowsky inventaron el hidrófono, un potente transductor ultrasónico con una frecuencia de resonancia de 150 KHz. En 1916 fue hundido el primer submarino detectado por un hidrófono. Sergei Y. Sokolov propuso en 1928 el uso de ultrasonidos para detectar fallas en metales.

A partir de 1938 Donald Griffin trabajó veinte años con murciélagos. Primero trabajó con J.W. Pierce quien había construido un micrófono especializado y un circuito para convertir los ultrasonidos en sonidos audibles. Encontraron que estos mamíferos nocturnos emiten ultrasonidos continuamente y más en vuelo. Griffin comprobó así la habilidad de ecolocalización ultrasónica en murciélagos.

Aplicación temprana del ultrasonido en la medicina

Langevin y otros encontraron que los ultrasonidos de alta intensidad tienen un poder destructivo. En un medio acuoso podían inducir dolor en una mano humana e incluso matar peces. Este tipo de ultrasonido se utilizó para la ablación de tejidos en el cerebro: ganglios basales en pacientes de Parkinson y lóbulo frontal en pacientes moribundos de cáncer. La piel, los huesos, el

tejido conectivo y los vasos sanguíneos son afectados mucho menos que el tejido nervioso y las células sensoriales. El ultrasonido se utilizó ampliamente en rehabilitación y también para tratar el mal de Menière. En los años 40 algunos intentaron hacer del ultrasonido una panacea.

A finales de los años 30 se experimentó con ultrasonido para el diagnóstico. Karl Theo Dussik intentó localizar tumores y los ventrículos en el cerebro. Dussik nombró su procedimiento hiperfonografía. Se trataba de un proceso complejo que requería sumergir la cabeza del paciente y apenas generaba imágenes como mosaicos. Wolf-Dieter Keidel experimentó con mediciones torácicas y cardíacas, también con un método de transmisión. Keidel solo obtuvo variaciones de la intensidad en relación con las pulsaciones cardíacas. André Dénier también probó sin éxito otro método de transmisión: ultrasonoscopia.

Expansión mundial de los métodos por reflexión

En 1949 se publica una parte del trabajo del médico naval George D. Ludwig, quien exploró sistemáticamente las características de los ultrasonidos en tejidos animales varios. Ludwig diseñó experimentos para detectar la presencia y localización de cuerpos extraños en tejidos animales incluyendo cálculos biliares. Él empleó un método de reflexión similar a la ecolocalización y determinó la velocidad del sonido en tejidos blandos entre 1490 y 1610 m/s, así como las frecuencias de escaneo óptimas entre 1 y 2.5 MHz. En 1950 se hace comercial el primer localizador ultrasónico para uso medicinal y biológico. En ese mismo año el cirujano John Wild y el ingeniero Donald Neal encuentran que el tejido intestinal maligno es más ecogénico que el benigno.

En 1952 Wild y John Reid publican su trabajo "Aplicación de técnicas de ecolocalización para determinar estructuras en tejidos biológicos". El año siguiente logran las primeras imágenes en tiempo real de crecimientos cancerosos en mama. Utilizaron un aparato modo B a 15 MHz para obtener imágenes de ultrasonido por modulación de intensidad en corte transversal. Trabajaron en detección de tumores de colon e inventaron aparatos modo A transvaginales y transrectales, pero se quedaron sin fondos. En 1956 Wild publicó un artículo sobre 117 nódulos de mama y reportó una precisión diagnóstica de más de 90%. John Reid trabajó luego en ecocardiografía con el especialista Claude Joyner.

También fue entre 1952 y 1953 que Douglas Howry y el nefrólogo Joseph Holmes trabajaron con un sistema en tanque de inmersión con el primer aparato modo B bidimensional. El sistema requirió diseño pionero en circuitos electrónicos como generador de pulsos, amplificación logarítmica, demodulación, y compensación ganancia tiempo. Así demostraron convincentemente la capacidad del ultrasonido para obtener imágenes de órganos internos y patologías. Luego Holmes trabajó con William Wright y Ralph Meyerdirk en un nuevo escáner que no requiriera un tanque de inmersión. En 1963 lanzaron a la venta un escáner de mano con brazo articulado y con base en un aparato modo B de contacto compuesto. Luego Holmes y sus colegas publicaron algunos de los primeros artículos en ultrasonido para gineco-obstetricia.

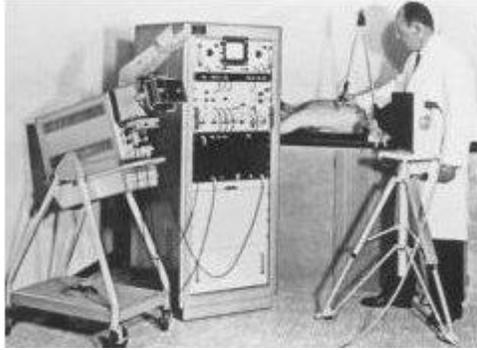


Figura 1: Escáner con brazo articulado [2].

En Japón colaboraron ingenieros, cirujanos y empresarios para producir un escáner ultrasónico modo A. Lo hicieron modificando un detector de fallas en metales. En 1952 publicaron 5 artículos sobre diagnóstico ultrasónico para enfermedades del cerebro. En 1955 fabricaron un escáner modo B con un arco de movimiento lineal y experimentaron con otros órganos. Dos años más tarde demostraron el uso de un aparato modo B con indicador de posición como en un RADAR. En ese mismo año Murooka y Wagai detectan cáncer de cérvix y agrandamiento uterino. En 1960 sacan a la venta un escáner modo B con bolsa de agua.



Figura 2: Escáner con bolsa de agua [2].

Reino Unido, Suecia y la Unión Soviética también desarrollaron el ultrasonido para diagnóstico médico. Tom Brown inventó y fabricó en 1957 el primer modo B compuesto con indicador de posición. Sunden reportó importantes hallazgos luego de investigar 400 casos de patologías pélvicas. Los soviéticos reportaron en 1962 diagnósticos de tumores compactos y fluidos en órganos genitales femeninos, endometriosis, piosalpinx y mioma.

La Primera Conferencia Internacional en Diagnóstico Ultrasónico tuvo lugar en Pennsylvania en 1965. El Primer Congreso Mundial sobre Diagnóstico Ultrasónico en Medicina tuvo lugar en Vienna en 1969 y el segundo en Rotterdam en 1972. También en 1972 se reunieron en Suiza delegados de 13 sociedades europeas de ultrasonido y fundaron la Federación Europea de Sociedades de Ultrasonido en Medicina y Biología. En 1977 se conforma la Sociedad Británica de Ultrasonido Médico.

Desarrollo del US en ginecología y obstetricia

A mediados de los años 60 se incrementó y expandió de manera notable la aplicación del ultrasonido en ginecología y obstetricia. Siemens en Alemania ofreció el primer escáner rápido modo B. En Austria surgió la compañía KretzTechnik que ofreció aparatos modo A y modo B. Kratochwil fue un investigador austriaco muy prolífico quien demostró la visualización de folículos ováricos en 1972. En Dinamarca también se reunió un equipo de trabajo y fundaron Smith Kline Instrument para fabricar en E.U.A. un aparato de su propio diseño que luego se vendió en España. En Finlandia se creó un centro de investigación utilizando aparatos estadounidenses. En Francia e Italia importaron aparatos austriacos. Los belgas probaron la tecnología japonesa. Polonia, los Países Bajos y China también experimentaron con los primeros aparatos de ultrasonido para el diagnóstico médico.

El grupo de trabajo de la Universidad de Juntendo reportó en 1963 el primer escaneo de un saco gestacional con 6 semanas de crecimiento. EL grupo fue muy activo y presentó trabajos en diagnóstico temprano, placentometría y encefalometría. En Wuhan, China, en 1963, reportaron hallazgos sonográficos en 261 embarazos anormales y en 1964 describieron ecocardiografía fetal modo M. Lamentablemente el trabajo se detuvo con la Revolución Cultural en 1966. En 1968, desde Sidney, Australia, publican Anatomía Fetal Desplegada por Ultrasonido y así empieza el diagnóstico de malformaciones fetales por ultrasonido. El uso efectivo de los escáneres se convirtió en un arte manual.

Stuart Campbell, de Glasgow, publicó en 1968 Un Método Mejorado para Cefalometría Fetal por Ultrasonido describiendo la maniobra de escáneres A y B para medir el diámetro biparietal fetal. Esta práctica se hizo un estándar para evaluar el crecimiento del feto y duró diez años. Muchos artículos se produjeron por esos años en cefalometría: Boog en Francia, Khentov en la Unión Soviética, Zacutti y Brugnoli en Italy, Kratochwil en Austria, Pystynen e Ylostalo en Finlandia. En 1973 Hugh Robinson, también de Glasgow, describió la medición de la longitud céfalo-caudal o cervex-coxis para estimar la edad gestacional. Campbell también reportó diagnósticos de anencefalia y spina bífida en 1972 y 1975.

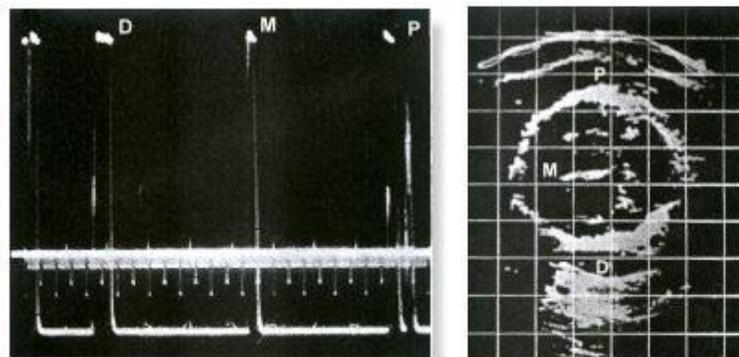


Figura 3: Cefalometría de Campbell. El modo A (izquierda) permitía hacer mediciones precisas y el modo B (derecha) permitía visualizar [3].

A principios de los años 80 existían 45 fabricantes de aparatos para ultrasonografía médica. En 1985 se comercializó por primera vez una sonda transvaginal y para 1988 la mayoría de los fabricantes tenían una opción endovaginal instalada en sus escáneres. El escaneo transvaginal se convirtió en una práctica estándar para manejar casos de infertilidad porque permitía evaluar el desarrollo folicular y el crecimiento del endometrio y también ayudaba en la extracción de ovocitos.

Más avances técnicos e inicio de la digitalización

El equipo de Georges Kossof reportó en 1971 en Melbourne escaneo obstétrico en escala de gris. Demostraron que los ecos de interés clínico provienen de la textura interna de los tejidos blandos. Optimizaron el procesamiento de señales para desplegar esta información de textura como niveles de gris. Una invención importante para la ecografía y el procesamiento de imágenes en general fue el convertidor de escaneo análogo. Una matriz de semiconductores era escaneada por un haz de electrones y, previo procesamiento, podía mostrarse en una televisión estándar, grabarse, escalar, aplicar pinzas y mediciones en pantalla.

Muchos intentaron aplicar un arreglo lineal de transductores sin verdadero éxito. El primer aparato de este tipo que se hizo comercial fue lanzado por ADR en E.U.A. en 1973 y utilizaba 64 cristales con técnicas de pasos [3]. En 1975 ADR mejoró su aparato implementando técnicas de enfoque. En 1978 Hitachi sacó a la venta el primer escáner con un convertidor digital. En ese mismo año los australianos reportaron sus resultados en imágenes de 512 por 512 píxeles y hasta 32 niveles de gris. En 1979 salió a la venta el primer escáner portátil.

El desarrollo de la tecnología de cómputo fue un factor importante en la mejora del desempeño en los aparatos de ultrasonido. La computadora digital DEC PDP-11 fue utilizada en esa época en diferentes aparatos de ultrasonido. Su principal sucesor fue el Motorola 6800. Otros factores fueron el aumento del número de transductores, el desarrollo de los cristales, el mejoramiento de los algoritmos, y el reemplazo de señales análogas por digitales.

Ultrasonido Doppler y alta resolución

En 1967 se delineó el uso de ultrasonido Doppler en obstetricia para la detección de vida fetal, localización de la placenta, flujo sanguíneo, y movimiento fetal. Al siguiente año Japón reportó las primeras mediciones Doppler de flujo en arteria umbilical. Fue hasta 1977 que se reportaron este tipo de mediciones en occidente. En 1980, los noruegos Eik-Nes y Marsal inventaron un escáner de arreglo lineal acoplado a un Doppler y documentaron velocidades de flujo en aorta fetal. Los australianos Giles y Trudinger estudiaron las formas de onda (de velocidad) y las utilizaron para cuantificar la tasa sistólica-diastólica. Publicaron tablas con los valores normales para diferentes fases gestacionales y así la velocimetría de flujo Doppler se hizo una práctica popular para evaluar el bienestar fetal.

En 1977 el francés Leandre Pourcelot describió las primeras imágenes codificadas en color Doppler. Fue hasta 1985 que los japoneses Namekawa y Kasai utilizaron detección de fase y una técnica de auto-correlación para producir imágenes de flujo en vasos sanguíneos. Este principio de mapeo de flujo por color se sigue utilizando hoy día. El primer artículo en evaluación fetal mediante dicho mapeo fue publicado por el croata Kurjak en 1987 utilizando un aparato japonés marca Aloka.

Las verdaderas mejoras en la calidad de imagen por ultrasonido se obtuvieron en los años 90 [4]. Fue gracias a los avances en tecnología de radar, telecomunicaciones, computación y televisión que la electrónica de alta velocidad se hizo accesible de manera económica. La cadena de procesamiento de señales se hizo totalmente digital. Se aumentó el ancho de banda y el número de canales en los transductores para mejorar la definición de las texturas. El desarrollo de arreglos bidimensionales ayudó a afinar el foco. La detección de fase permitió aumentar la resolución y la tasa cuadros por segundo. El desarrollo de transductores con excelente linealidad en la transmisión y con una sensibilidad superior permitieron la obtención de imágenes por armónicos. Esta técnica permite usar frecuencias bajas para mejorar la penetración mientras que solo detecta y procesa los ecos de mayor frecuencia que caracterizan los tejidos.

2.2 Ultrasonido y cerebelo fetal

Se puede elegir un transductor de acuerdo con el tipo físico de la madre y con la posición del feto. En general, se puede hacer un examen básico con transductores transabdominales y frecuencias de 3 a 5 MHz. Muchas veces es necesario un transductor transvaginal con frecuencias de 5 a 10 MHz. La ecografía tridimensional y el uso de armónicos pueden mejorar el detalle anatómico y facilitar la evaluación del sistema nervioso central (SNC).

La edad gestacional

La configuración del cerebro y de la columna vertebral evoluciona durante la gestación [5]. Para evitar errores diagnósticos, es importante reconocer las apariencias normales del SNC en las diferentes edades gestacionales. La mayoría de los trabajos para diagnosticar anomalías neuronales se centran a la mitad de la gestación. Los exámenes básicos se realizan generalmente alrededor de las 20 semanas de gestación. Algunas anomalías pueden ser visibles desde el primer trimestre. Éstas pueden representar una minoría, pero por lo general son graves. La evaluación temprana requiere habilidades especiales. La ventaja de un examen temprano entre las 14 y 16 semanas es lo delgado de los huesos. Entonces el cerebro puede ser evaluado desde casi todos los ángulos.

La determinación de la edad gestacional es muy importante en el control del embarazo [6]. Cuando este parámetro no es correcto se tiene riesgo de parto prematuro y de embarazo prolongado. Gracias al ultrasonido se ha disminuido de manera significativa la edad gestacional dudosa. En el primer trimestre se mide saco gestacional y longitud céfalo-nalgas. En el segundo trimestre se mide diámetro biparietal y longitud del fémur. Cuando hay sospecha de feto grande para edad gestacional o de retardo del crecimiento intrauterino se mide el diámetro transversal

máximo del cerebelo. Esta medición requiere una imagen de la fosa posterior incluyendo la cisterna magna, el cerebelo completo, y el pliegue nucal.

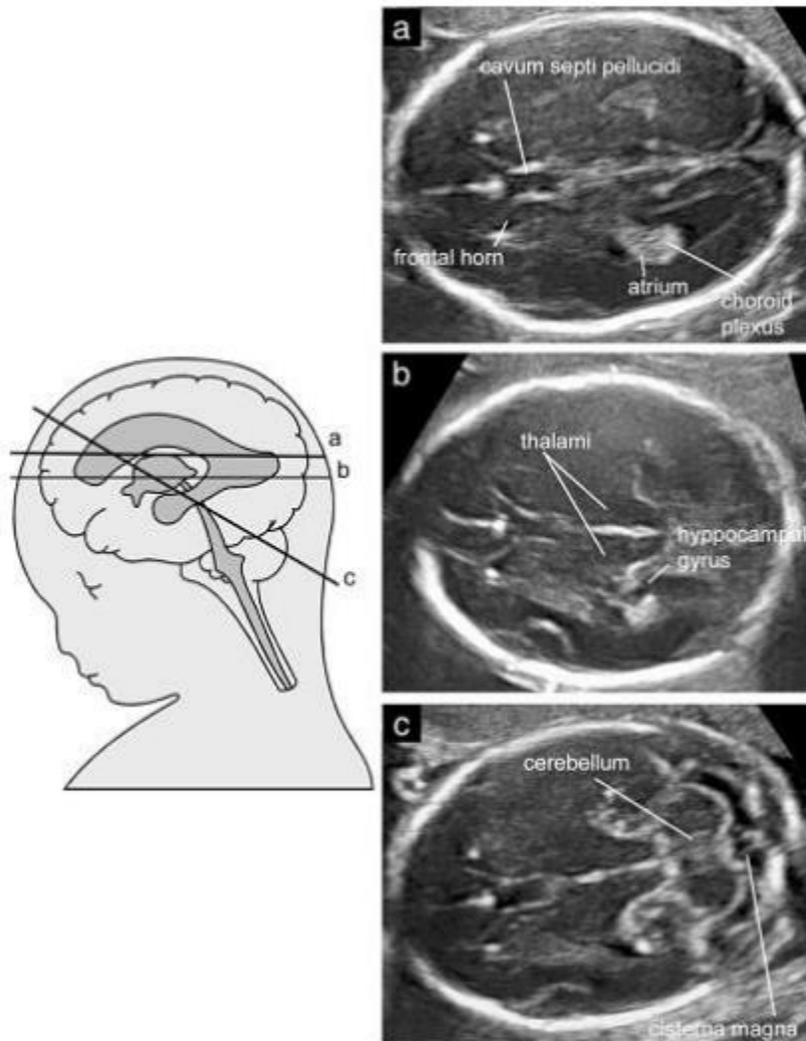


Figura 4: Vistas axiales de la cabeza fetal. (a) Plano transventricular; (b) plano transtalámico; (c) plano transcerebelar [5].

Evaluación del sistema nervioso central

Las malformaciones del sistema nervioso central (SNC) son algunas de las más comunes entre todas las anomalías congénitas. El ultrasonido se ha utilizado durante al menos 30 años para ayudar a diagnosticar anomalías del SNC fetal. Un examen básico del cerebro fetal es el reconocimiento de la anatomía. La evaluación detallada del SNC fetal también es posible, pero requiere de una habilidad especial y de una tecnología sofisticada. Este examen avanzado está indicado en los embarazos con un mayor riesgo de anomalías del SNC.

Aproximadamente uno de treinta mil nacimientos presenta el síndrome Dandy-Walker y éste implica un desarrollo anormal o ausencia del cerebelo [7]. Otra anomalía igualmente rara se llama Arnold Chiari e incluye alteraciones en la forma del cerebelo. Estas malformaciones difícilmente se detectan antes del segundo trimestre.



Figura 5: Vista transversal de cabeza fetal (semana 26) demostrando encefalocele. La herniación de meninges y cerebelo son características del síndrome Arnold-Chiari III [7].

Diferentes campos de estudio han aportado evidencia para señalar que el cerebelo tiene funciones motoras, cognitivas, y afectivas [8]. El cerebelo tiene conexiones con zonas motoras y no motoras. Una lesión o degeneración del cerebelo presenta un síndrome motor, cognitivo, y afectivo. Una teoría propone que el cerebelo realiza un cómputo uniforme que modula tanto procesos motores como cognitivos y afectivos. Las similitudes entre anomalías motoras y lingüísticas en pacientes con enfermedades cerebelosas refuerzan esta idea.

2.3 Inteligencia Artificial y Ciencias Cognitivas

Orígenes

Las raíces más antiguas de la Inteligencia Artificial pueden atribuirse a dos ideas de Gottfried Leibniz (filósofo, matemático y político) en el siglo XVII [9]. *Characteristica universalis* sería un supuesto conjunto de pensamientos primitivos capaces de abarcar todo el razonamiento humano, y *calculus ratiocinator*, un conjunto de reglas útiles para deducir mecánicamente todas las verdades posibles a partir de dichos pensamientos. Son la semántica y la sintaxis de un lenguaje ideal que permitiría razonar sin ambigüedades [10]. Una realización parcial de dicho lenguaje podría caracterizar a una máquina inteligente, así como el lenguaje natural caracteriza al humano.

Podemos seguir estas raíces hasta el siglo XIX con dos obras sobre lógica: *System of Logic* de Stuart Mill (filósofo y político) y *Laws of thought* de George Boole (filósofo y matemático). La primera trata de incorporar y sistematizar las mejores ideas que se hubieran escrito sobre el tema hasta entonces [11]. La segunda presenta un sistema de reglas que antecedió a la lógica como una ciencia formal que hoy forma parte de la filosofía y de las matemáticas. Siendo el pensamiento la actividad característica de la inteligencia, entonces sus leyes tienen que gobernar, conformar o estar al origen de la inteligencia artificial.

Formalización y primeros trabajos

En 1950 Alan Turing publica *Computing machinery and intelligence*, donde describe una prueba que permitiría determinar si una computadora se puede considerar inteligente. Propone que un humano y una máquina interactúen un tiempo utilizando lenguaje natural, si el humano no puede reconocer que está tratando con una máquina entonces la máquina es inteligente. En 1955 tiene lugar un congreso de verano que propone investigar sobre el problema de la inteligencia artificial suponiendo que cada aspecto de la inteligencia puede describirse tan precisamente que sea factible una máquina capaz de simularlos. Proponen siete líneas de investigación [12]:

- Programación y computación eficientes
- Traducción entre lenguaje natural y de máquina
- Redes neuronales
- Estimación del costo computacional
- Desarrollar esquemas de auto mejora para máquinas inteligentes
- Clasificar las diferentes formas de abstracción y describir sus métodos
- Aleatoriedad e intuición para simular pensamiento creativo

Ese mismo año Herbert Simon (economista y politólogo), J. C. Shaw (informático) y Allen Newell (informático) crearon *Logic Theorist*, y al año siguiente, *General Problem Solver*, dos de los primeros programas que trabajaban con razonamiento simbólico y que fueron considerados inteligentes, pero no de acuerdo con la prueba de Turing.

Una tendencia importante en los años 1960 fue la traducción automática gracias al interés del ejército de EE. UU. en traducir documentos del ruso. Parece que varios investigadores desestimaron la complejidad de la tarea o fueron demasiado optimistas con sus predicciones. Al pasar de los años no alcanzaban los resultados esperados y en 1966 se recortaron todos los fondos para dichas líneas de investigación.

Las ciencias cognitivas como respuesta a la IA

En Reino Unido se comisionó a Sir James Lighthill (matemático) para reportar objetivamente el estado de la investigación en IA. El reporte fue presentado en 1973 y apoyó la investigación relativa a la simulación por computadora de procesos neurofisiológicos y psicológicos, pero criticó fuertemente áreas como robótica y procesamiento del lenguaje [13]. Lighthill agrupó la

investigación sobre IA en tres ramos: ramo A de *automation* (incluyendo reconocimiento de texto y discurso, traducción, búsqueda de información, diseño y manufactura, análisis, solución de problemas, y toma de decisiones), ramo C de *central nervous system* (se refiere al cerebro, a la conducta, y al aprendizaje; teoriza e interpreta datos de neurobiología y psicología; construye modelos computacionales para desarrollar la teoría), y ramo B de *bridge* y de *building robots* (es la investigación y la actividad que reúne o conecta los ramos A y C con el objetivo de construir robots).

Longuet-Higgins (químico), en su respuesta al reporte, menciona por primera vez a las ciencias cognitivas [14]. Sugiere que todas las ciencias relevantes para el pensamiento humano pueden enriquecerse con los estudios sobre IA y propone cuatro cabeceras: matemáticas, lingüística, psicología, y fisiología. En 1975 Bobrow y Collins publican el libro *Cognitive Science* y dos años después empieza un periódico homónimo. Las ciencias cognitivas se definen por el conjunto de problemas que ataca y el conjunto de disciplinas que reúne. A diferencia de las cabeceras que son analíticas y de la inteligencia artificial que es sintética, las ciencias cognitivas persiguen un equilibrio entre estas dos formas.

De acuerdo con la clasificación de la ACM (*Association for Computing Machinery*) de 2012, la IA y el Aprendizaje de Máquina son dos ramas diferentes dentro de las metodologías de la computación. La IA se divide en ocho sub-ramas: procesamiento del lenguaje natural, razonamiento y representación del conocimiento, planificación y programación, metodologías de búsqueda, métodos de control, fundamentos teóricos y filosóficos, sistemas distribuidos, y visión computacional.

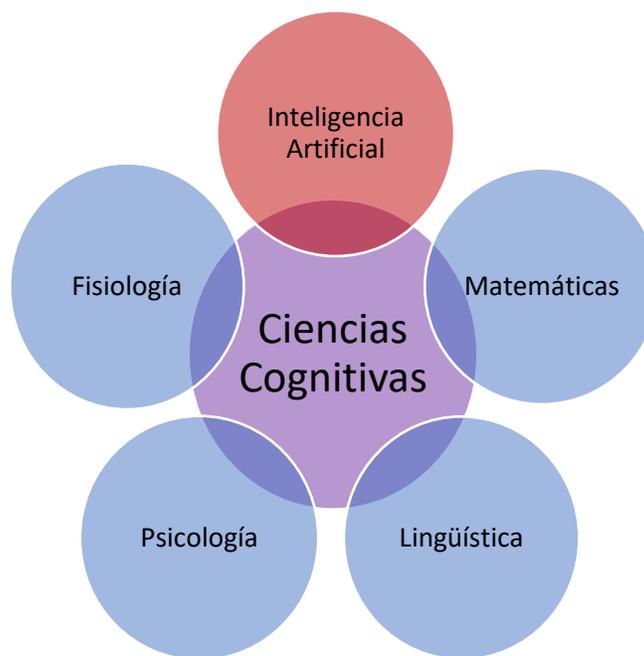


Figura 6: Ciencias Cognitivas e Inteligencia Artificial

2.4 Redes Neuronales Artificiales y Aprendizaje Profundo

Dinámica neuronal y optimización de sistemas

El origen de las Redes Neuronales Artificiales lo podemos encontrar en dos obras de los años 1940 [15]. El artículo de Walter Pitts (investigador) y Warren McCulloch (neurofisiólogo), *A Logical Calculus of Ideas Immanent in Nervous Activity*, presenta la idea de las redes neuronales artificiales: “y para la biofísica matemática, la teoría aporta una herramienta para el tratamiento simbólico riguroso de redes conocidas y un método fácil de construir redes hipotéticas de propiedades requeridas” [16]. El libro de Hebb (psicólogo), *The Organization of Behavior*, da origen a la psicología fisiológica que investiga la relación entre la biología y procesos como el aprendizaje y la memoria [17], reúne elementos de fisiología y de psicologías clínica, animal y experimental para formar una teoría general de la conducta [18].

Frank Rosenblatt (psicólogo) publica en 1958 su trabajo temprano con el perceptrón [19], hoy en día, el modelo básico de la neurona artificial. La preocupación de Frank no era la invención de un dispositivo para la inteligencia artificial sino investigar las estructuras físicas y los principios neuro-dinámicos que subyacen a la inteligencia natural [20]. En su libro de 1962, *Principles of Neurodynamics: perceptrons and the theory of brain mechanisms*, y en 1967, con su trabajo titulado *Recent work on theoretical models of biological memory*, exploró las ideas de redes multicapas y de la propagación hacia atrás del error, que son claramente precursoras del Aprendizaje Profundo.

En 1974 Paul Werbos (economista) presenta su tesis doctoral *Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences*, donde investiga la aplicación de la estadística y la econometría para predicción política. Dicha obra dedica un capítulo a “la retroalimentación dinámica, estimación estadística y la optimización de sistemas”, donde desarrolla la propagación hacia atrás del error. El año siguiente Fukushima publicó su trabajo para la interpretación de caracteres escritos a mano: *Cognitron: A self-organizing multilayered neural network*. La premisa del Cognitrón es que la sinapsis de la neurona ‘x’ a la neurona ‘y’ es reforzada cuando ‘x’ dispara y ninguna neurona en la vecindad de ‘y’ dispara más fuerte que ‘y’ [21].

Stephen Grossberg (matemático) construye un modelo para clasificación de patrones utilizando mecanismos (como memorias de corto y largo plazo) encontrados experimentalmente en el córtex visual de gatitos. En 1976 demuestra que estos modelos pueden reajustarse para responder a cualquier conjunto de patrones espaciales predefinidos. De esta forma una red neuronal permite transformar cualquier conjunto de características en cualquier otro conjunto si tiene suficientes neuronas [22].

Conexionismo y cómputo en paralelo

En 1981 Hinton y Andrews publicaron *Parallel models of associative memory*. El libro es una colección de artículos presentados en una conferencia informal dos años atrás [23]. La edición incluye representantes de neurofisiología, psicología cognitiva, ciencia de cómputo y matemáticas.

Estos puntos de vista se ponen de acuerdo en algo: “interacciones paralelas entre unidades simples proveen una mejor metáfora de la actividad mental que largas secuencias de pasos simples”.

Kohonen propone en 1982 un principio para facilitar la formación automática de mapas de características a partir de eventos observables. Su objetivo primordial era de mostrar que los estímulos externos por sí solos pueden promover estructura y funcionalidad en un sistema [24]. En el mismo año Hopfield presenta *Neural Networks and Physical Systems with Emergent Collective Computational Abilities*, trabajo que describe redes neuronales recurrentes funcionando como un sistema de memoria accesible por contenido.

En 1985 el *American Institute of Physics* inaugura las reuniones anuales sobre redes neuronales para computación. En 1987, el IEEE celebra la primera conferencia abierta sobre el tema y conforma la INNS (*International Neural Network Society*). Carpenter y Grossberg proponen, ese mismo año, una máquina para reconocimiento de patrones. Describen una arquitectura masiva de neuronas auto organizada en forma paralela. En 1988 la INNS estrena su periódico *Neural Networks*, al año siguiente el MIT inicia *Neural Computation*, y en 1990 el IEEE comienza con *Transactions on Neural Networks*.

Redes neuronales profundas

A finales de los años 1990 se llevaron a cabo trabajos fundamentales para el origen del aprendizaje profundo. De acuerdo con el Descubridor de Información de la UNAM [25], el primer artículo en mencionar redes neuronales profundas es *Deep Feedforward Networks: Application to Pattern Recognition*, publicado por Babri y Tong en 1996 [26]. El año siguiente Hochreiter y Schmidhuber resolvieron un problema con las redes recurrentes que impedía aprender a almacenar información sobre periodos de tiempo extendidos. En su artículo *Long short-term memory*, introducen unidades especiales llamadas compuertas ya que abren o cierran el acceso al flujo constante del error [27]. En 1998 LeCun y varios produjeron LeNet-5, la primera red convolucional, y obtuvieron resultados significativos en reconocimiento de caracteres escritos a mano.

Según diversos autores [28] [29], Aizenberg hace la primera mención del término “aprendizaje profundo” en el año 2000. Aizenberg se refiere a los trabajos realizados con lógica de umbrales (threshold logic) en los años 1960: “un aprendizaje profundo de características de funciones booleanas de umbral” [30]. Esencialmente la lógica de umbrales utilizó neuronas artificiales para representar funciones lógicas binarias [31]. La lógica de umbrales consiguió que el modelo de neurona artificial llegara a la electrónica de consumo en forma de compuerta lógica avanzada y se utilizó en la fabricación de algunas computadoras digitales [32] pero esto no corresponde con el concepto de aprendizaje profundo ligado a las redes neuronales profundas.

En 2006 Hinton y otros publican *A fast learning algorithm for deep belief nets*, mejorando los resultados obtenidos por LeCun en 1998. Andrew NG indica que fue Geoffrey Hinton quien acuñó el aprendizaje profundo como “la nueva generación de redes neuronales” [33] y Hinton dio una plática homónima en 2007 [34]. Dicha plática incluyó temas como el entrenamiento de redes

profundas y los autocodificadores profundos [35]. La teoría de complejidad en circuitos sugería que una red profunda debía requerir menos elementos computacionales para representar funciones con un alto nivel de abstracción, pero no se encontraba la forma de entrenar o ajustar sus parámetros [36]. Hinton desarrolló uno de los primeros métodos para entrenar redes profundas haciendo una optimización previa y no supervisada capa por capa. Y así empezó el aprendizaje profundo.

3. Análisis automático de imágenes médicas

El aprendizaje profundo se aplica en ultrasonido médico para clasificación, segmentación, detección, registro, biometría, intervención guiada, y terapia [37]. Las tareas más básicas son clasificación, detección, y segmentación. Estas tareas básicas se han aplicado en diversas estructuras anatómicas como mama [38], [39], próstata [40]–[42], hígado [43], corazón [44], [45], cerebro [46], [47], arteria carótida [48], [49], tiroides [50], vasos sanguíneos [51], feto [52]–[57], nodos linfáticos [58], riñón [59], columna vertebral [60], huesos [61], [62], músculos [63], y nervios [64].

3.1 Detección

La detección de objetos de interés como son tumores y nódulos es esencial para el análisis en imágenes de US. La detección y localización de estos objetos pueden servir como punto de apoyo para proceder con la clasificación y la segmentación.

- Cáncer en próstata

En 2015 Azizi et al. [41] publicaron una propuesta de un marco de trabajo para análisis frecuencial de señales ultrasónicas en tejido de próstata. Utilizaron una red *deep belief network* (DBN) para extraer características a partir de las componentes espectrales. Realizaron un primer entrenamiento no supervisado y un segundo entrenamiento supervisado bajo los datos de histopatología. Luego emplearon una máquina de soporte vectorial (SVM) para estimar la probabilidad de cáncer a partir de las características. Con este marco de trabajo obtuvieron un área bajo la curva ROC de 0.91.

- Lesiones de mama

Cao et al. [65] realizaron un comparativo entre cuatro modelos de aprendizaje profundo para detección de objetos aplicados en la detección de lesiones de mama. Los modelos comparados fueron Regional CNN (R-CNN), Faster R-CNN, You Only Look Once (YOLO), y Single Shot Multi-Box Detector (SSD). SSD obtuvo los mejores resultados en precisión y exhaustividad.

- Planos estándar (US fetal)

La obtención de un plano estándar en ultrasonido fetal para la evaluación del embarazo requiere un alto grado de expertise. La detección de malformaciones y las estimaciones biométricas dependen de la obtención correcta de un plano estándar. Baumgartner et al. [66] han logrado en 2016 la detección automática de doce de vistas estándar relativas a: riñones, cerebro, abdomen, columna, fémur, y corazón. El método utiliza una red totalmente convolucional y podría asistir a operadores que todavía no tienen el expertise necesario.

- Fase del ciclo cardíaco

La medición de diferentes parámetros cardíacos se apoya en la caracterización de la fase del ciclo cardíaco a partir de datos de ecocardiografía. Dezaki et al. [67] combinaron redes

residuales (ResNets) con redes recurrentes (RNN) para extraer jerarquías de características (cuadro por cuadro) y modelar dependencias temporales (entre cuadros). Así lograron un método para el reconocimiento automático de la fase en el ciclo cardiaco.

- Regiones de interés (nervio)

La anestesia regional guiada por ultrasonido ha tomado relevancia en los últimos años [64]. Un sistema automático para la segmentación del nervio podría ayudar al practicante a concentrarse en la administración del anestésico. Hafiane et al. publicaron en 2017 una propuesta que empleó CNN para la detección y localización de regiones de interés. Luego aplicaron *consistencia espaciotemporal* para obtener una detección robusta y finalmente delinearon la región de interés mediante un modelo de contornos activos.

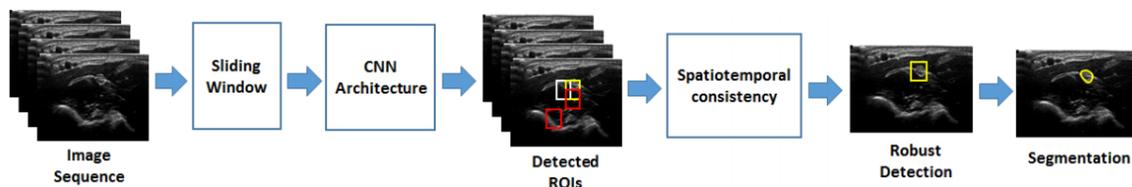


Figura 3.1: Redes neuronales convolucionales para la detección de regiones de interés dentro de un flujo de procesamiento para la segmentación de nervio.

3.2 Clasificación

Es una tarea fundamental en radiología diagnóstica que consiste en discriminar diferentes estructuras anatómicas o tipos de tejidos. No existe una inteligencia artificial que emule la cadena de razonamiento completa necesaria para la interpretación de una imagen médica. La investigación en sistemas de diagnóstico asistido por computadora (CADx) ha concentrado esfuerzos en la clasificación automática de objetos de interés como nódulos, tumores, y lesiones.

- Tumores de mama

El cáncer de mama es la segunda causa de muerte en mujeres. La principal herramienta para su diagnóstico es la mamografía, pero el US también es utilizado para el examen de mama y la clasificación de tumores. En 2012 Jamieson et al. [68] aplicaron una red neuronal profunda para el aprendizaje no supervisado de características en US de mama. Luego utilizaron aprendizaje supervisado para clasificar lesiones benignas y malignas a partir de las características producidas por la red profunda. Obtuvieron resultados tan buenos como si hubieran extraído las características mediante un diseño tradicional. Varios trabajos siguieron con la aplicación de aprendizaje profundo para el diagnóstico de tumores de mama. En 2016 y 2017 los trabajos de Cheng JZ [69] (utilizando *stacked denoising auto-encoders*) y Han S [70] (utilizando *GoogLeNet*) obtuvieron resultados que superaban el desempeño de los humanos expertos.

- Cáncer de hígado

El cáncer de hígado es la tercera causa de muerte por cáncer en el mundo [71]. Un diagnóstico temprano y preciso es importante para aumentar la tasa de supervivencia. En 2014 Wu et al. [43] aplicaron *Deep belief networks* (DBN) sobre curvas de intensidad-tiempo en secuencias de ultrasonido mejorado en contraste. Obtuvieron 86.4% de precisión superando métodos convencionales de aprendizaje de máquina.

- Nódulos tiroideos

Los nódulos en la tiroides son una de las lesiones nodulares más comunes en la población adulta y su diagnóstico requería una biopsia. El ultrasonido es una herramienta de diagnóstico no invasiva y se puede beneficiar de un sistema automatizado para la clasificación de nódulos. En 2017 Ma et al. [72] fusionaron dos arquitecturas CNN, una profunda y una superficial, para hacer clasificación de nódulos en la tiroides. El mismo año Liu et al. [73] fusionaron una CNN profunda con mapas de características diseñados, también para clasificación de nódulos, y obtuvieron una mejora de 14% sobre del empleo solo de mapas de diseño.

- Calidad de imagen

En examen de ultrasonido prenatal incluye la medición de la circunferencia abdominal (CA). El bajo contraste y lo irregular de las formas dificultan la tarea y pueden causar un error en la estimación de la edad gestacional. Por esto es importante la evaluación de la calidad en imágenes de ultrasonido fetal. Wu et al. publicaron en 2017 su propuesta de un sistema de evaluación de la calidad en imagen de US fetal [74]. Dicho sistema tiene dos etapas, una CNN para localizar regiones de interés, y otra CNN para clasificar (en buena o mala calidad) el plano estándar que se ocupa en la medición de la CA.

- Planos estándar (corazón fetal)

Localizar el corazón fetal e identificar el plano en una imagen de US fetal son tareas importantes para la detección de enfermedades cardíacas congénitas. Sundaresan et al. [75] emplearon una red totalmente convolucional (FCN) para segmentar el corazón y así obtener directamente la localización y la clasificación en cuatro planos estándares.

3.3 Segmentación

Es el proceso que separa una imagen en un número de regiones uniformemente homogéneas [76]. Cada región es una parte constitutiva o un objeto de la escena entera. La segmentación de una imagen es un conjunto de regiones conectadas que no se traslapan. Así cada píxel o punto en la imagen pertenece a una región. La segmentación es muy importante en el análisis automático de imágenes porque la obtención de objetos o regiones permite continuar con procesos como la descripción o el reconocimiento.

La segmentación de estructuras anatómicas es un paso previo al análisis cuantitativo de parámetros clínicos relativos al volumen y a la forma. También es útil para la generación de regiones de interés que permitan procesos de análisis y clasificación subsecuentes. La segmentación precisa de estructuras anatómicas y lesiones es un desafío por las características de las imágenes por US.

- Ventrículo izquierdo (corazón)

La ecocardiografía es una de las modalidades más comunes para visualizar y diagnosticar patologías en el ventrículo izquierdo (VI). La segmentación de dicho ventrículo en momentos específicos ayuda al cardiólogo en su análisis. La automatización de dicha tarea se dificulta por las variaciones en la forma, el ruido, las sombras, y lo discontinuo de los bordes. Cuando se trata de VI fetal la tarea se complica aún más debido a los movimientos del feto. Yu et al. [77] publicaron en 2016 un método para la segmentación de VI fetal con CNN dinámicas y afinación (fine-tuning). El método requiere que el primer cuadro de la secuencia se anote manualmente y con este ejemplo se hace la afinación de todas las capas de la red (afinación profunda). Con la red afinada y el segundo cuadro se realiza la segmentación y ésta se toma como ejemplo para afinar las últimas capas (afinación superficial). Se sigue aplicando la afinación superficial con el resto de la secuencia. El método también incluye un seguimiento de los puntos base de la válvula mitral para la separación del ventrículo y del atrio.

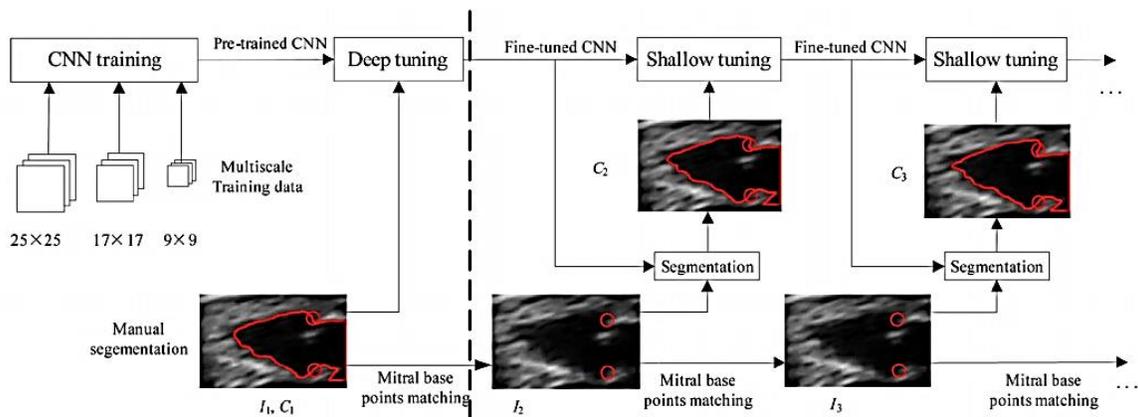


Figura 7: Segmentación de VI fetal con CNN dinámicas [77].

- Nódulos linfáticos

La imagen por ultrasonido de nódulos linfáticos se ha utilizado ampliamente para el diagnóstico clínico [58]. Una segmentación precisa del nódulo permite los análisis cuantitativo y morfológico que pueden proveer información importante para un diagnóstico. Zhang et al. propusieron redes totalmente convolucionales apiladas de grueso a fino (CFS-FCN). Primero entrenaron una red FCN para segmentar nódulos y objetos parecidos. Para esto utilizaron anotaciones no expertas. Luego combinaron las segmentaciones producidas por esta red con

las imágenes originales. Esta información combinada y las anotaciones expertas fueron utilizadas para entrenar una segunda red. Su método trabajó de lo grueso a lo fino, de lo simple a lo complejo, y obtuvo los mejores resultados para esta tarea en su momento.



Figura 8: Modelo apilado de FCN que produce un resultado intermedio con el objeto buscado y otros similares [58].

- Hueso (vértebras)

En cirugía de hernia lumbar hay que localizar el nivel correcto en las vértebras y esta no es una tarea trivial [78]. Con rayos X preoperatorios se delimitan y etiquetan las vértebras. En el quirófano se utiliza US para escanear la columna. Baka et al. probaron U-Net para la segmentación de las vértebras estableciendo un nuevo estado del arte en 2017. Su propuesta también incluyó el emparejamiento con las radiografías para propagar las etiquetas con el nivel de vértebra a la imagen de US.

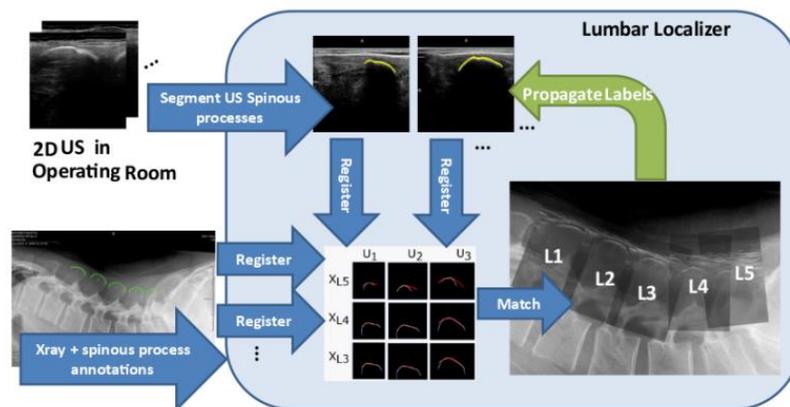


Figura 9: Segmentación de hueso con aprendizaje profundo. Registro, emparejamiento y propagación de información preoperatoria a imagen por US en el quirófano [78].

- Próstata

La segmentación automática de la próstata en US se dificulta por discontinuidades en el perímetro, bordes ambiguos, y una severa heterogeneidad [42]. Yang et al. realizaron un marco de trabajo que integra extracción de características (aprendizaje profundo) e información a priori de la forma (ampliamente usada en métodos tradicionales). Su método estima el perímetro de la próstata con un enfoque secuencial. Primero convierten la imagen en

secuencias para estimar posibles perímetros de acuerdo con las formas a priori. Luego fusionan las diferentes estimaciones para obtener un mapa probabilístico. Finalmente integran de forma iterativa el último mapa obtenido con la imagen original (autocontexto) para afinar la estimación del perímetro.

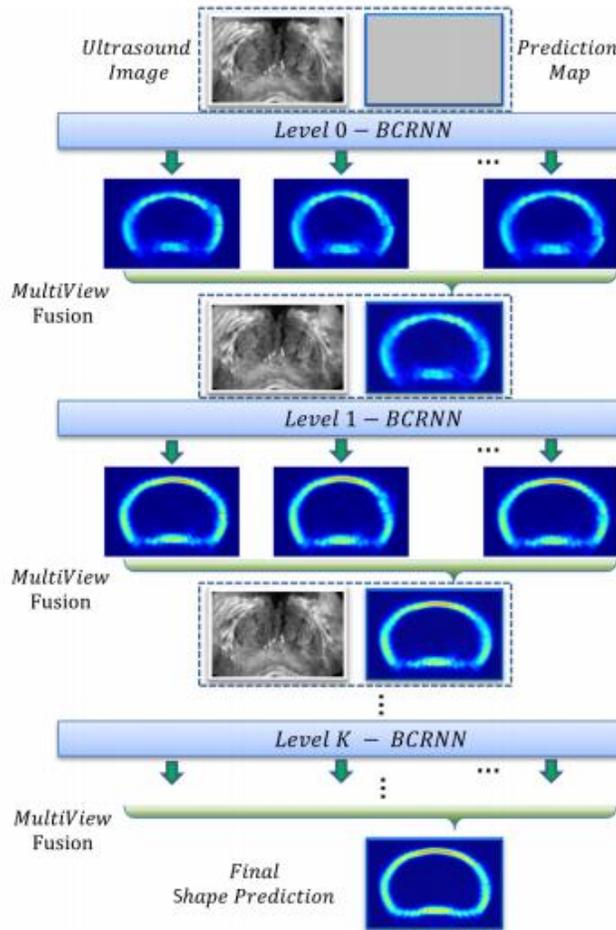


Figura 10: La estimación de la forma se refina de forma gradual mediante RNN para completar bordes (BCRNN) [42].

- Ultrasonido prenatal

Las discontinuo y ambiguo de los bordes en imágenes de US también obstaculiza la delineación o segmentación automática de abdomen y cerebro fetal para el análisis biométrico [79]. Wu et al. implantaron una versión aligerada de FCN-8 en un esquema de autocontexto que llamaron *cascaded FCN* o casFCN. Dicho esquema suma la imagen de entrada con la última predicción o estimación del borde (en forma de mapa probabilístico) para integrar el contexto local. Con cada nivel de autocontexto la estimación se afina y mejora el resultado final en la segmentación.

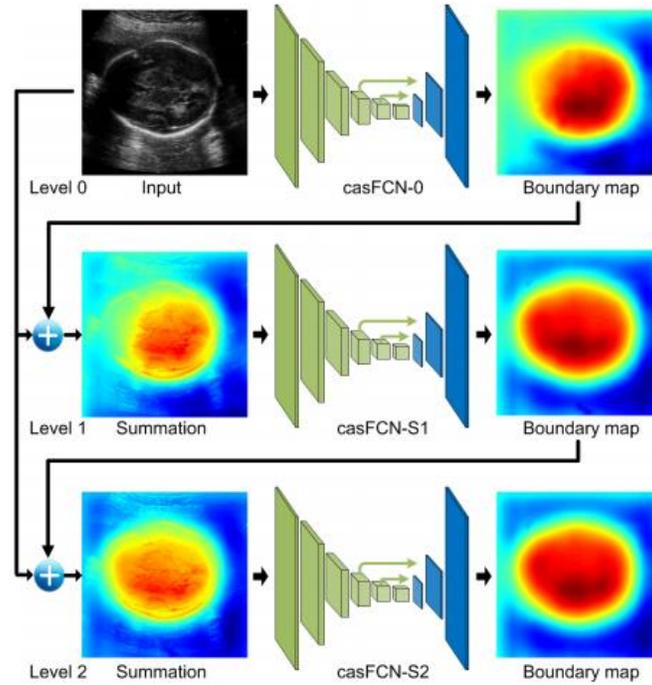


Figura 11: Implantación de FCN en cascada mostrando dos niveles de autocontexto [79].

3.4 Análisis 3D

- Endometrio

La medición del grosor en el endometrio ayuda a la evaluación de rutina para pacientes con un problema de fertilidad y al examen de cáncer para mujeres que presentan sangrado postmenopausia [80]. Singhal et al. presentaron en 2017 un método para medir automáticamente el grosor del endometrio con imágenes de US tridimensional. Su algoritmo combina aprendizaje profundo con un método más interpretable, un modelo variacional para segmentación mediante propagación de curvas. Aplica FCN (U-Net) y utiliza el mapa de probabilidad de endometrio como funcional de energía para la segmentación.

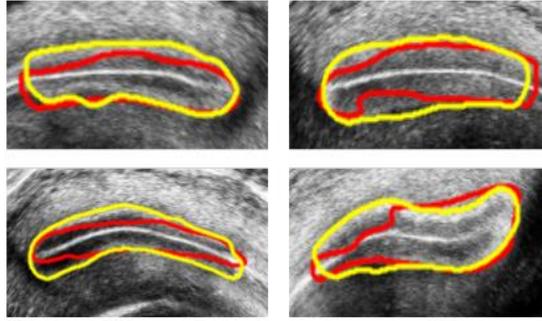


Figura 12: Segmentaciones de endometrio con U-Net y U-Net más propagación de curvas, en rojo y amarillo respectivamente [80].

- Epidural guiada por US

El objetivo para una inyección epidural puede localizarse de manera automática con US para evitar posibles errores y complicaciones [81]. Un método híbrido con aprendizaje de máquina y aprendizaje profundo fue propuesto por Pesteie et al. en 2018. Primero clasifica los planos en el volumen de US mediante características locales direccionales de Hadamard y una red neuronal densa (FC). Los planos clasificados como plano objetivo pasan a un proceso de localización por parches. Se aplican de forma paralela CNN y Hadamard. Ambas características pasan por una red densa y finalmente se aplica una capa Softmax.

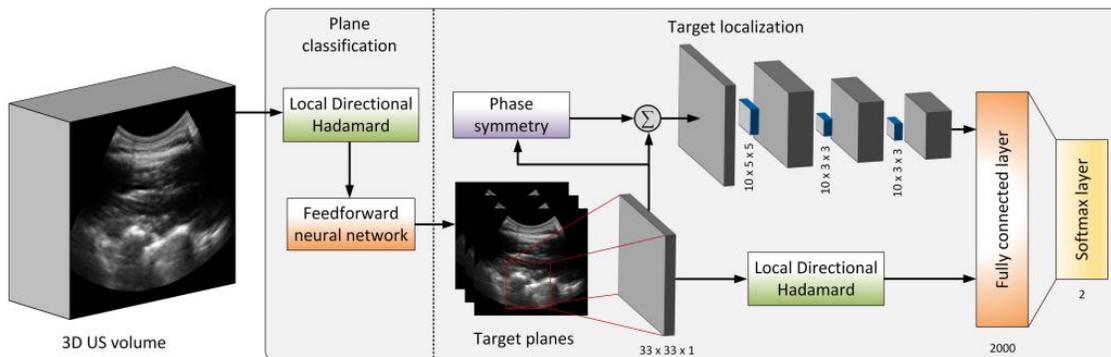


Figura 13: Características de Hadamard y aprendizaje profundo para tareas de clasificación y localización [81].

- Feto, saco gestacional, y placenta

Los métodos 3D que se han visto realizan, en realidad, un trabajo 2D plano por plano. Requieren menos memoria y pueden apoyarse en modelos 2D previamente entrenados. Un verdadero trabajo 3D no cuenta con estas ventajas, pero si puede incorporar la información contextual entre planos. Yang et al. [82] implementaron segmentación 3D automática en volúmenes de US fetal. Primero implementaron FCN 3D para la clasificación de cada punto

(voxel) en el volumen. Luego aplicaron LSTM para aprender las relaciones espaciales de forma secuencial y refinar los resultados o reducir el error en los volúmenes serializados.

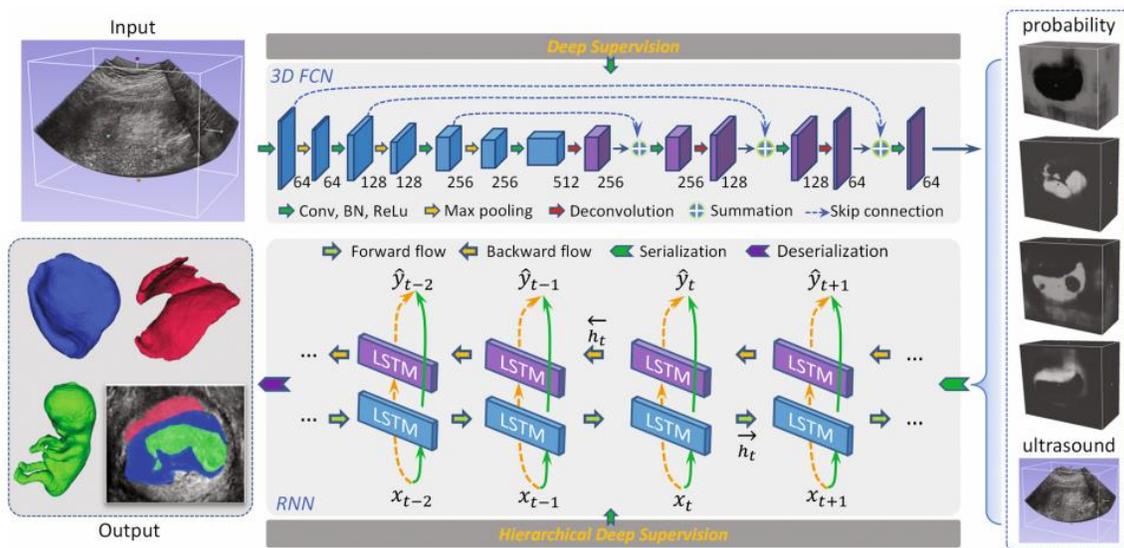


Figura 14: Aplicación de FCN 3D y RNN para segmentación de volúmenes [82].

4. Discusión

En esta sección se reflexiona sobre los alcances de esta obra y se propone una nueva forma de aplicar aprendizaje profundo en imágenes médicas por US. Se resumen algunas relaciones entre disciplinas que fundamentan este trabajo y se expresan de forma gráfica. Se reconoce que la comunidad y el registro de su actividad en el ciberespacio son parte fundamental de este trabajo. Presentamos un breve análisis sobre del estado del arte presentado en el capítulo 3 y mencionamos algunas aplicaciones de aprendizaje profundo fuera de imágenes médicas para poner en contexto dicho estado del arte.

4.1 Alcances

Hay que reconocer que se trabajó con imágenes entregadas por un aparato moderno y manipulado por una persona experta. También fue de gran ayuda contar con anotaciones realizadas por una especialista. Un estudio exhaustivo tendría que revisar las técnicas de manipulación y anotación, el procesamiento digital de señales (PDS) desde los transductores, la tecnología en los transductores, y la caracterización de los tejidos. Desafiando los límites descritos y considerando la capacidad de las redes profundas nos planteamos la posibilidad de trasladar la parte del procesamiento de señales para la formación de la imagen a un modelo de aprendizaje profundo.

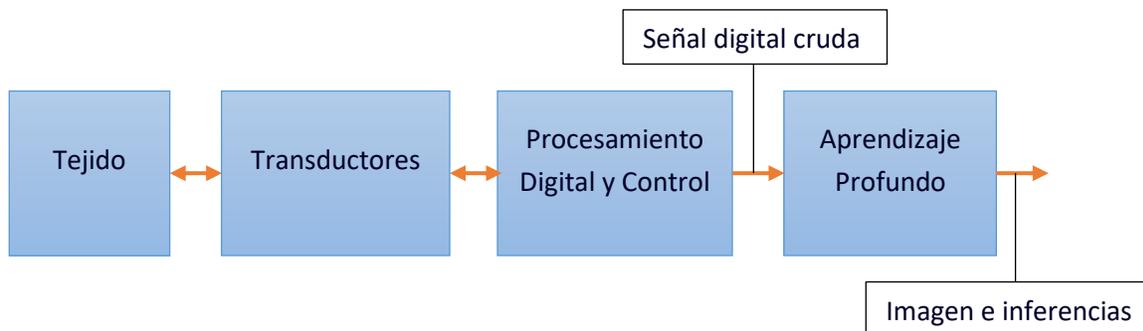


Figura 15: Adquisición y procesamiento de señales ultrasónicas en un sistema CADx.

4.2 Antecedentes

La inteligencia artificial (IA) y el aprendizaje de máquina (AM) se pueden considerar ramas separadas dentro de las metodologías de la computación. Aprendizaje profundo (AP) es un tipo de aprendizaje de máquina que está avanzando el estado del arte en ramas de IA como son búsqueda de la información, procesamiento del lenguaje natural y visión computacional. Entonces AP establece un puente o una conexión entre la IA y el AM. IA es una rama en computación y también está al origen de las ciencias cognitivas. AM y AP también son formas de procesamiento digital de señales. Procesamiento digital de señales y aprendizaje profundo tienen muy diversas aplicaciones

y en esta tesis se trabajó con imágenes médicas. Las relaciones recién descritas se pueden expresar en un diagrama de Venn para visualizar los fundamentos establecidos en el capítulo 2.

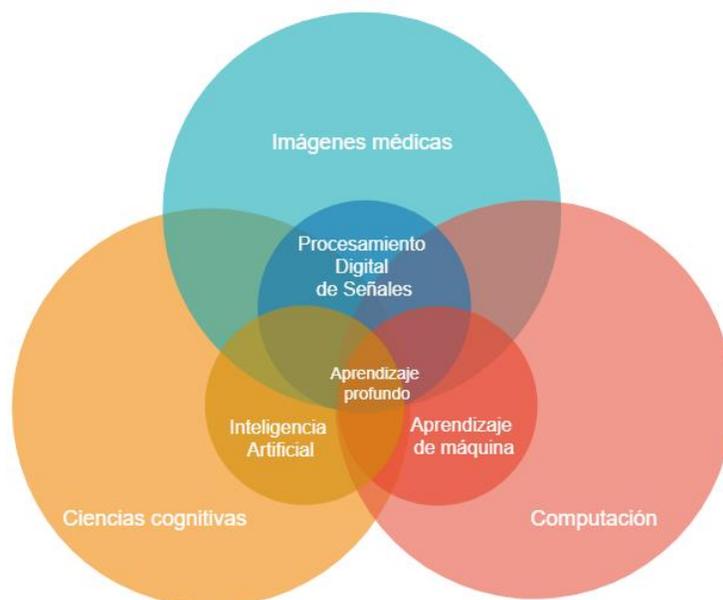


Figura 16: Visión amplia de los fundamentos para el análisis de imágenes médicas mediante aprendizaje profundo.

Por lo que se vio en los orígenes y desarrollos de US médico, inteligencia artificial y aprendizaje profundo, comprobamos que la ciencia y el arte se construyen a partir de las interacciones entre personas. Los patrones de construcción más recurrentes en esta obra han sido las publicaciones, los congresos, y las conferencias, pero también sabemos de los seminarios, talleres, tutorías, pláticas informales, retos públicos y redes sociales. Gracias a la comunidad que trabaja en las fuentes de código abiertas es posible descargar de internet diferentes marcos de trabajo para aprendizaje de máquina, aprendizaje profundo y más. También está la comunidad que participa en la documentación del código y en los comentarios sobre la implementación.

4.3 Estado del arte

Utilizamos un sistema de revisión semiautomática [83] para procesar y analizar el conjunto de 44 referencias formado por nuestra revisión o resumen del estado del arte. Dicho sistema nos reporta las referencias por año y por tipo. Observamos un patrón de crecimiento importante entre 2014, 2015, 2016, y 2017 con 1, 3, 10, y 25 publicaciones respectivamente. Esto significaría prácticamente una triplicación de publicaciones cada año. El reporte contó 19 artículos, 2 libros, y 23 conferencias. Finalmente obtenemos una visualización con las palabras claves más relevantes.



Figura 17: Palabras clave más relevantes en nuestro resumen del estado del arte.

Hacemos una revisión manual y contamos las referencias para diferentes modelos de aprendizaje profundo: 6 de CNN, 5 de FCN, 2 de U-Net, 2 de DBN, y 2 de RNN. Otras redes que se mencionan una sola vez: ResNet, AutoEncoder, y LSTM. Métodos tradicionales de aprendizaje automático y procesamiento de imágenes que se utilizan junto con aprendizaje profundo: ventana deslizante, análisis multi escala, máquina de soporte vectorial, modelo de contorno activo, propagación de curvas y transformada de Hadamard. Se observa que es bastante común la concatenación de modelos.

Otras aplicaciones

Aprendizaje profundo ha reunido muchos éxitos en pocos años en una variedad de aplicaciones. Muchos experimentos han demostrado un desempeño superior sobre aprendizaje automático tradicional en áreas como procesamiento de imágenes, visión computacional, procesamiento de lenguaje natural, robótica y control, bioinformática, ciberseguridad, y más [84].

- *Procesamiento de lenguaje natural*

En los últimos años se aplicó aprendizaje profundo en una diversidad de tareas relacionadas con el lenguaje natural. Las tareas más básicas son etiquetado gramatical y análisis sintáctico [85]. También se han implementado tareas como clasificación de texto, extracción de relaciones semánticas, y análisis de sentimiento. En el siguiente nivel de dificultad se pueden considerar traducción automática, respuestas a preguntas, y resumen de textos. Finalmente hay que

mencionar las tareas más complejas como son reconocimiento del discurso (voz), descripción de imágenes, y descripción de escenas (video) [86].

- *Ciberseguridad*

Tareas de ciberseguridad que se han realizado aplicando redes neuronales profundas son: análisis y detección de malware en PC y Android, detección de intrusiones, fraudes (phishing), correo basura, y desfiguración de sitios web [87].

5. Metodología

Los modelos conexionistas o arquitecturas de redes neuronales se construyen a partir de hiperparámetros como son número de capas, número de filtros y tamaño de filtro (para capas convolucionales), reducción de dimensiones (max-pooling), tipo de activación y tamaño de la salida (capas densas). Los parámetros o pesos y sesgos producen mapas característicos o conjuntos de valores a partir de la imagen en la entrada (el ejemplo de aprendizaje) o la capa anterior. Este proceso se propaga hacia adelante hasta la salida donde se compara con la salida esperada y se mide el error. Los parámetros se optimizan o se ajustan para reducir el error desde la salida hacia la entrada (propagación hacia atrás) capa por capa. Una vez que el error es pequeño se procede a evaluar con datos “nuevos” para la red o ejemplos no introducidos anteriormente. Por esto se hace una partición de datos para aprendizaje o entrenamiento y otra para prueba o evaluación.

5.1 Datos y métricas

Se obtuvieron 21 volúmenes de ultrasonido fetal gracias a la colaboración entre la UNAM y el Instituto Nacional de Perinatología (INPER) [88]. De cada volumen se extrajeron 15 imágenes con el plano transcerebelar y 14 planos paralelos (7 inferiores y 7 superiores). Una especialista del INPER anotó cada una de las 315 imágenes delineando el contorno del cerebelo. A partir de las anotaciones se generaron máscaras de segmentación binarias (blanco y negro) en archivos binarios de Matlab. Así es que se cuenta con 315 imágenes de US en escala de gris y sus respectivas máscaras de segmentación para el presente trabajo. El tamaño promedio de las imágenes es de 255 * 173 píxeles.

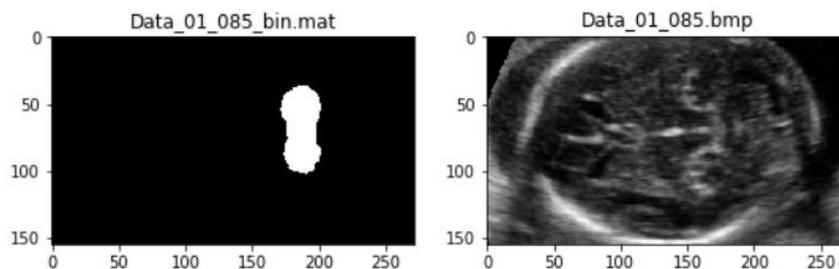


Figura 18: Ejemplo de máscara de segmentación con el cerebelo en blanco y su imagen de US correspondiente.

El puntaje F1 evalúa la detección aplicando la media armónica entre precisión y sensibilidad. La precisión es el número de objetos detectados correctamente entre el número de objetos presentes. La sensibilidad es el número de objetos detectados correctamente entre el número total de objetos detectados.

$$F1 = 2 \cdot \frac{\text{precisión} \cdot \text{sensibilidad}}{\text{precisión} + \text{sensibilidad}} \quad (1)$$

La intersección sobre la unión (IoU) o índice de Jaccard (J) evalúa la localización comparando los rectángulos delimitadores (el anotado y el inferido). También evalúa la segmentación comparando el área inferida y la anotada, pixel por pixel. Se calcula J para cada imagen de prueba y luego se obtiene el promedio o precisión promedio (AP) para las imágenes de prueba.

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|} \quad (2)$$

5.2 Descenso por gradiente

El aprendizaje en la red empieza con la carga de la arquitectura en la memoria y aplicando valores aleatorios en los pesos cuando no se tienen valores ya entrenados. Los pesos representan la fuerza de las uniones entre las neuronas artificiales. Se carga un ejemplo en la entrada y se propaga la información hacia adelante hasta obtener una inferencia en la salida. Entonces se calcula el error entre la salida esperada (la anotación correspondiente al ejemplo) y la salida inferida. Luego se calcula el gradiente del error con respecto a los pesos $\nabla_w E(w_t)$. El algoritmo de gradiente descendiente más simple modifica sus pesos a cada paso o ejemplo con la multiplicación del gradiente por la tasa de aprendizaje ϵ [89]:

$$\Delta w_t = -\epsilon \nabla_w E(w_t) \quad (4)$$

El descenso por gradiente con momento se acelera sumando el coeficiente de momento μ por la modificación de los pesos en el paso anterior Δw_{t-1} . La desventaja es que se deben conservar en memoria dichos elementos del paso anterior.

$$\Delta w_t = -\epsilon \nabla_w E(w_t) + \mu \Delta w_{t-1} \quad (5)$$

5.3 Detección y localización

Arquitectura YOLO

YOLO es una arquitectura de aprendizaje profundo para detección, clasificación y localización de objetos en imágenes [90]. Es el acrónimo en inglés de la frase "solo se mira una vez". Sus autores la compararon con el método de aprendizaje de máquina llamado Modelo de Partes Deformables (DPM) [91], y también con Redes Neuronales Convolucionales Regionales (R-CNN) [92]. Los tres diferentes métodos fueron evaluados de acuerdo con el reto PASCAL 2007. YOLO resultó cientos de veces más rápido que los otros dos métodos. La rapidez en la inferencia es su principal característica. Resultó el doble de preciso que DPM y casi igual a R-CNN. YOLO fue innovador porque resolvió un problema de regresión para predecir la localización (coordenadas y tamaños) de marcos delimitadores y las clases asociadas. La segunda característica de la red YOLO es la detección unificada. Los autores la describen como un razonamiento global de la imagen y sus objetos. Explican que esto se logra con la división de la imagen en una cuadrícula con predicciones separadas para cada celda. La tercera característica es su capacidad para generalizar entre dominios y lo comprobaron usando imágenes artísticas. La arquitectura se compone de 24 capas convolucionales

(para la extracción de características) y dos capas totalmente conectadas (para la detección, clasificación y localización de objetos).

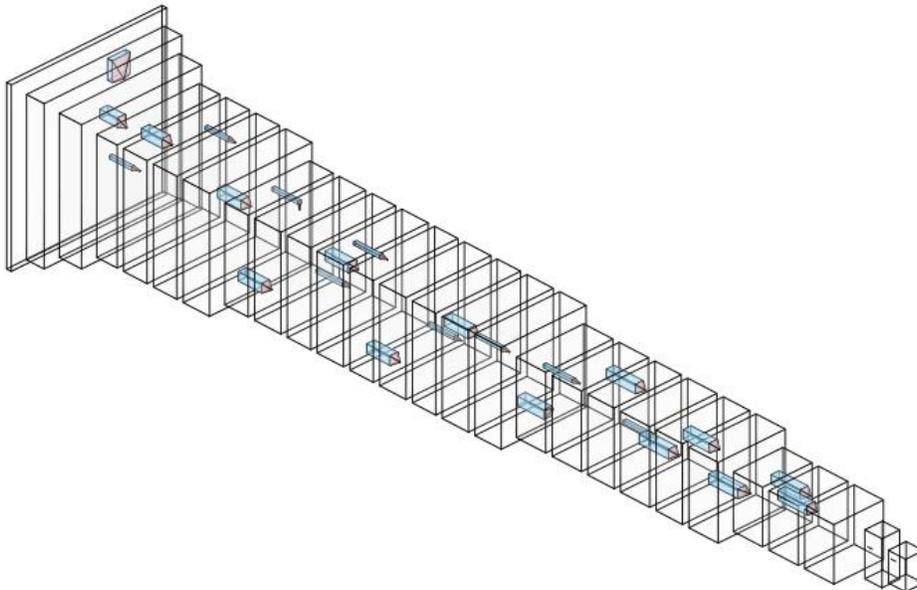


Figura 19: Arquitectura YOLO con 24 capas convolucionales.

Anotaciones y partición

El entrenamiento de YOLO requiere un archivo de texto con anotaciones para cada imagen de ejemplo. Cada anotación debe incluir 5 valores: número de clase, coordenadas (x,y), ancho y alto del rectángulo delimitador. Cuando una imagen contiene diferentes objetos se hace una anotación para cada uno en renglones subsecuentes. En este caso tenemos una sola clase y un solo objeto por imagen: el cerebelo. Se escribió un guión para transformar los archivos binarios de Matlab en archivos de texto, las máscaras en anotaciones YOLO. Un segundo programa fue necesario para generar archivos de texto con las listas de archivos para entrenamiento y para pruebas. Esta partición de datos se realizó de forma aleatoria asignando aproximadamente el 75% de imágenes al aprendizaje y el 25% restante a la evaluación. Los guiones fueron interpretados por Python [93].

Marco de trabajo

Se implementó la red en una computadora equipada con procesador x86 de 4 núcleos y frecuencia de 3.5 GHz, con 24 GB de memoria volátil, y con 4GB de memoria en una GPU. El código fuente es una biblioteca llamada Darknet [94], está disponible para Linux y se compiló en Ubuntu 18.04. Darknet tiene mejor rendimiento utilizando CUDA [95] y una GPU compatible. Este rendimiento “mejorado” es necesario para ejecutar el entrenamiento o aprendizaje con la red YOLO. CUDA es un conjunto de herramientas de programación para el procesamiento en paralelo con GPUs

de la marca NVIDIA. Nuestra GPU es compatible y cuenta con 1344 núcleos para procesamiento en paralelo.

Darknet incluye la implementación de YOLO para PASCAL-VOC [96] con 20 clases de objetos. Para aprender una sola clase (cerebelo) se modificó la fuente y se recompiló el archivo ejecutable. Las imágenes en escala de gris se transformaron al mapa de color *viridis* de Matplotlib [97] para cumplir con un requerimiento de la red que está hecha para trabajar con imágenes en color RGB. Dicho mapa va de un azul oscuro al amarillo, pasando por un verde, y presenta uniformidad perceptual y robustez a la ceguera de colores [98]. La uniformidad perceptual asegura que no se alteren los cambios de intensidad de gris al pasar la imagen a una paleta de colores.

Experimentos

Para probar la hipótesis A entrenamos YOLO con nuestros datos de cerebelo. Las configuraciones de la red y del entrenamiento se describen en un archivo de texto. Este archivo también se modificó para ajustar la salida de la red a una sola clase. La arquitectura termina en un tensor de salida con un número B de predicciones para cada celda en la imagen de entrada. Cada predicción incluye las coordenadas y las dimensiones del rectángulo delimitador y la probabilidad de tener un objeto (de cualquier clase). Además, se predice, en cada celda, la probabilidad de clase para un número C de clases. La imagen de entrada se divide o analiza en S celdas por lado. Entonces el número de elementos en el tensor de salida es:

$$S * S * (B * 5 + C) \quad (3)$$

En la siguiente figura veremos una representación gráfica del tensor de salida [99] con una cuadrícula de lado $S=7$, $B=2$ predicciones por celda (incluyendo coordenadas, dimensiones, y probabilidad de objeto), y $C=1$ clase probable.

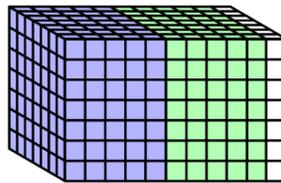


Figura 20: Representación gráfica del tensor de salida [99].

Para probar la hipótesis B se realizaron 4 experimentos utilizando diferentes valores para S: 7, 9, 11, y 13. Los respectivos números de elementos en el tensor de salida son 539, 891, 1331, y 1859. En todos los casos de hacen 2 predicciones por celda. El entrenamiento o aprendizaje de la red se limitó a de 1000 iteraciones con un tamaño de lote de 64 imágenes. La partición de entrenamiento tiene 237 imágenes. Si dividimos el total de imágenes procesadas entre el total disponible obtenemos el número de épocas: 270. Es el número de veces que la red ve o repasa la totalidad de los datos durante el aprendizaje.

Se implementó transferencia de aprendizaje utilizando pesos previamente entrenados con ImageNet [100]. El archivo de pesos se descargó desde la página web de Joseph Redmon [101], principal autor de YOLO. Los pesos se utilizaron para inicializar las capas convolucionales y esto significa que la transferencia de aprendizaje afectó únicamente la extracción de características. Las neuronas dedicadas a las tareas de detección y localización tuvieron que aprender “desde cero”. La optimización de los pesos se realizó mediante descenso por gradiente estocástico con una tasa base de aprendizaje $\epsilon=5 \cdot 10^{-4}$ y con un factor de momento $\mu=0.9$. La tasa de aprendizaje se redujo multiplicando por una escala hasta cierto número de iteraciones de acuerdo con la siguiente regla: hasta 200 iteraciones se multiplica por 2.5, hasta 400 por 2, hasta 600 por 1, hasta 800 por 0.5 y hasta 1000 por 0.25.

Tiny YOLO

Se trata de una arquitectura similar a YOLO, pero incluye 8 capas convolucionales en lugar de 24. Además de reducir el tamaño de la red, se redujo también el tamaño de entrada de $448 \cdot 448$ a $224 \cdot 224$ píxeles. Con estos cambios, el costo computacional y el tiempo de aprendizaje tienen una disminución significativa. Se condujeron 8 experimentos para probar la hipótesis C y para observar los efectos por la variación del factor de momento en el aprendizaje. Se ensayaron 4 valores para el factor de momento: 0.3, 0.5, 0.7, 0.9.

5.4 Segmentación semántica

Las redes totalmente convolucionales son el estado del arte en segmentación semántica de imágenes por US mediante aprendizaje profundo. En la sección 3.3 incluimos cinco trabajos, entre los cuales contamos 3 que utilizan FCN. Y podemos sumar los dos trabajos de la sección 3.4. Entonces se ha visto que, en nuestro análisis del estado del arte, el 71% de las tareas de segmentación ocupan FCN. Por esto se entrenó FCN con datos de cerebelo y para probar la hipótesis D.

Redes totalmente convolucionales

Fully Convolutional Networks (FCN) [102] o redes totalmente convolucionales demostraron que es posible entrenar una arquitectura compuesta de capas convolucionales y deconvolucionales para segmentación semántica. Tres versiones de FCN se construyeron a partir de la arquitectura VGG16 [103] para detección y clasificación de objetos en imágenes. Aumentaron el tamaño de los últimos mapas de características (conv7) y reemplazaron las capas densas por capas deconvolucionales para obtener FCN-32. Aumentaron más y aún más el tamaño de conv7 para sumar mapas de características menos profundos (pool4 y pool3) y añadieron capas deconvolucionales para obtener FCN-16 y FCN-8. En estas versiones se combinan diferentes niveles de información para obtener mejores resultados. Las capas deconvolucionales aprenden a redimensionar los mapas de características, aumentando su tamaño 8, 16, y 32 veces, hasta obtener mapas de segmentación con las dimensiones de la imagen de entrada. Se obtienen tantos mapas

probabilísticos como clases de objetos se aprenden. Al final se logra la clasificación píxel por píxel tomando el máximo entre clases probables.

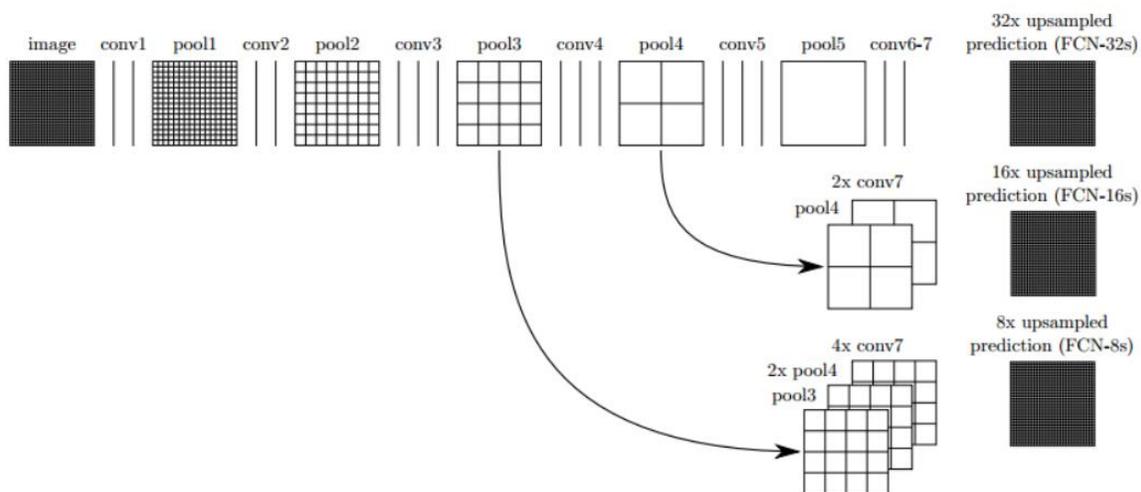


Figura 21: Primeras tres versiones de redes totalmente convolucionales para segmentación semántica [102].

Anotaciones y partición

Dada la tarea de segmentar o clasificar punto por punto, es necesario hacer las anotaciones de la misma manera. En este caso las anotaciones se corresponden con las máscaras de segmentación binarias. La clase 0 corresponde a lo que no es cerebro o al fondo y la clase 1 se corresponde al cerebro. Solo fue necesario convertir los archivos en formato Matlab (.mat) a formato Portable Network Graphics (.png). Puesto que los archivos .png deben tener 3 o 4 canales de información se triplica el único canal disponible. La partición de datos se realizó de forma aleatoria, tomando únicamente 10 archivos para prueba.

Marco de trabajo

El hardware utilizado para YOLO no fue suficiente para trabajar con FCN-8. Entonces se utilizó una computadora equipada con procesador x86 de 8 núcleos a una frecuencia de 3.6 GHz., 64 Gb de memoria volátil, y 3 GPUs (cada uno con 3584 núcleos y 11 Gb de memoria). Nuestra implementación se realizó con Keras [104], a partir de la fuente de Yumi [105]. Keras es una interfaz de programación para procesamiento de datos, prototipado rápido y experimentación con modelos conexionistas utilizando Tensorflow [106]. Este último se describe como una plataforma de aprendizaje de máquina y facilita la ejecución de operaciones con tensores y el cómputo de gradientes.

Inicialización de pesos

Glorot experimentó con redes densas de cinco capas y encontró que la rápida saturación (hacia cero) de las activaciones en las últimas capas resulta en un aprendizaje global muy

ralentizado [107]. Entonces sugiere utilizar una inicialización normalizada para evitar dichos estancamientos. Su método, conocido como Xavier o Glorot, considera los números de elementos en la entrada y en la salida (n_{in} y n_{out}) para asignar valores a los parámetros de la red:

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_{in}+n_{out}}}, \frac{\sqrt{6}}{\sqrt{n_{in}+n_{out}}}\right] \quad (6)$$

He experimentó con redes convolucionales y densas (incluyendo VGG) y estudió la inicialización con un enfoque en las activaciones no lineales [108]. Propuso un método propio y observó que tenía mejor desempeño comparado con Glorot. Experimentó con una red de 30 capas y encontró que el aprendizaje se estancaba completamente con el método Glorot mientras que su método sí logró hacer convergencia. Por lo tanto, He se considera apto para redes muy profundas:

$$W \sim U\left[-\frac{\sqrt{6}}{\sqrt{n_{in}}}, \frac{\sqrt{6}}{\sqrt{n_{in}}}\right] \quad (7)$$

Experimentos

Datos centrados y no centrados

Se generó un conjunto de datos con el cerebelo centrado a partir de las anotaciones, más un margen para incluir parte del contexto. Esto se hizo para probar la hipótesis E y para simular la ventaja que significaría tener la información de localización como pueden ser las estimaciones producidas por YOLO. Entonces contamos con dos conjuntos de datos: centrados y no centrados. En una primera serie de ensayos se probaron los 2 conjuntos de datos, 2 arquitecturas CNN para extracción de características: VGG16 y VGG19, 4 algoritmos para la optimización de los parámetros en la red, y 2 modos de entrenamiento: con transferencia de aprendizaje o desde cero. Estos elementos se pueden combinar de 32 formas diferentes, 32 es entonces el número de entrenamientos para esta primera serie. Los 4 algoritmos ensayados son descenso estocástico del gradiente (SGD), un método con tasa de aprendizaje adaptada (Adadelta) [109], un método para optimización estocástica (ADAM) [110], y ADAM con momento de Nesterov (Nadam) [111].

Datos aumentados por simetría

Se ha comentado que cuando no se tienen pesos previamente entrenados se aplican valores aleatorios, pero existen diferentes maneras de proponer estos números. Realizamos una segunda serie de experimentos para probar dos formas de inicializar los parámetros sin transferencia de aprendizaje. Estos elementos se encuentran en la última capa convolucional y en todas las capas deconvolucionales. Se prueban nuevamente las 2 versiones de VGG. Para trabajar con la hipótesis F se ensayan 2 modos de entrenamiento: con transferencia de aprendizaje y desde cero o totalmente aleatorio. Estos 8 ensayos se realizaron con un nuevo conjunto de datos que considera la apariencia simétrica del cerebelo: se conservaron las mitades superiores y las inferiores se reflejaron de forma vertical. De esta forma se duplicó el número de ejemplos para aprendizaje y entonces podemos decir que se realizó un aumento de datos por simetría. Las imágenes de medio cerebelo conservan su relación de aspecto original ya que la entrada de la red tiene forma cuadrada y no rectangular

(que ajusta mejor a la forma del cerebello entero). En la siguiente figura veremos una superposición de algunos contornos extraídos a partir de las máscaras de segmentación.

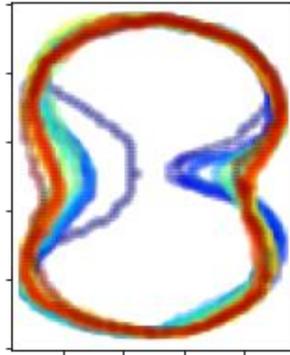


Figura 22: Contornos a partir de las anotaciones.

Datos clasificados

Se confeccionaron dos conjuntos más de datos para probar la hipótesis G. Por conceptos anatómicos estos conjuntos se pueden nombrar como proximal y distal. Las imágenes del conjunto anterior (mitad de cerebello) se dividieron con un eje vertical. Además, cada conjunto se agrupó en 18 clases de contornos caracterizados por estadísticos de sus coordenadas. En esta forma se pretende que la red aprenda a segmentar porciones de cerebello correspondientes a un cuadrante del original. Entrenamos con estos conjuntos de datos disminuyendo el tamaño de entrada a 160*160 píxeles. En la evaluación, las 18 clases de cerebello (por su contorno) se confunden en una misma clase: cerebello. En la siguiente figura veremos la superposición de contornos agrupados por estadísticos de las coordenadas.

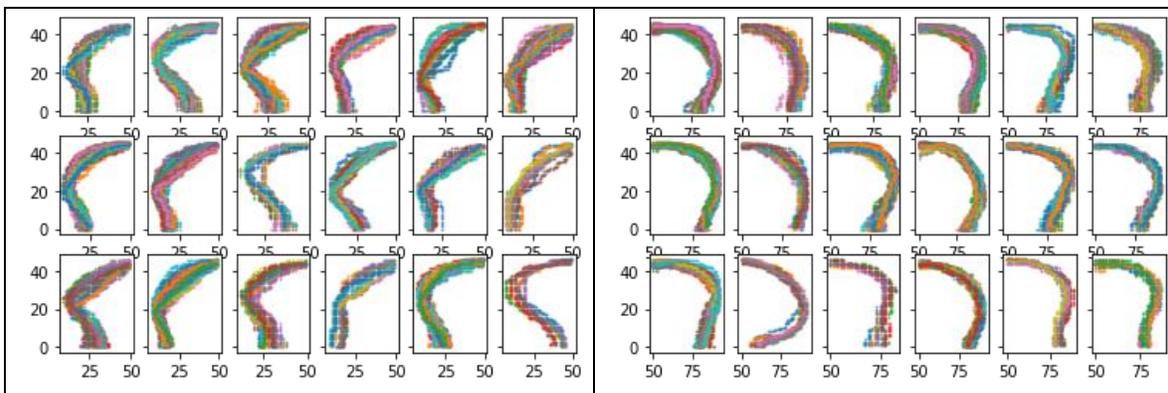


Figura 23: Grupos de cuartos de contorno proximal (izquierda) y distal (derecha).

6. Resultados

6.1 Detección y localización

YOLO

Se ejecutaron 4 entrenamientos de 1000 lotes. Cada 100 lotes se guardaba el conjunto de pesos o modelo aprendido hasta el momento. Así se obtuvieron 40 modelos para su evaluación en las tareas de localización y detección. Las métricas que se aplican respectivamente son la precisión promedio AP y el puntaje F1. La evaluación de los modelos obtenidos en diferentes momentos del entrenamiento nos muestra la evolución del aprendizaje. En la siguiente figura se muestra un gráfico para la evaluación de cada tarea y en cada gráfico la evolución del aprendizaje para 4 variantes de YOLO.

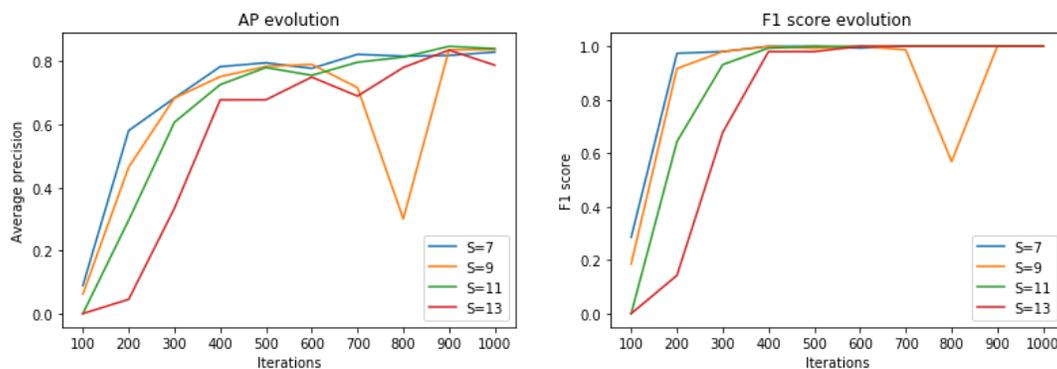


Figura 24: Evaluación de localización la izquierda (precisión promedio) y detección a la derecha (puntaje F1).

Recordamos que se realizaron 4 experimentos para diferentes valores del hiperparámetro S que es el número de divisiones por lado. Para cada experimento se toma la precisión promedio más alta y se anotan en la siguiente tabla refiriendo el momento del entrenamiento como el número de lotes procesado.

Divisiones	49	81	121	169
Lotes procesados	1000	1000	900	900
Precisión promedio	82.97 %	83.77%	84.8%	83.6%

Tabla 1: Mejores resultados con YOLO luego de evaluar 40 modelos.

Se observó entre los mejores resultados que el incremento en el hiperparámetro S permitió la mejora de la precisión promedio excepto para S=13. En la evolución de las métricas se puede ver que el incremento de S significó en general un aprendizaje más lento. ¿Podíamos esperar los mejores resultados con S=13 si hubiéramos extendido el entrenamiento?

Se vio que el número de divisiones se traduce en un mayor número de elementos en el tensor de salida y esto implica un mayor número de conexiones o costo computacional. ¿El aumento de las divisiones permite obtener mejor precisión o es el número de elementos en la última capa densa? Para despejar esta duda se propone realizar nuevos experimentos aumentando el número B de predicciones por celda y conservando constante el número S de divisiones.

Sus autores afirman que YOLO razona globalmente y de una vez toda la imagen, pero la redundancia en las predicciones se opone a esta idea. Se puede intuir que el número de divisiones confiere a la red una resolución en la predicción. O sea que esta cuadrícula tiene un rol en la formación de características útiles para la tarea de localización. Entonces nos preguntamos si la redundancia es todavía necesaria para la tarea de predicción en sí. Así surge la idea de tomar un modelo entrenado con muchas divisiones, congelar o hacer constantes los pesos en las capas convolucionales y entrenar nuevas capas densas para una salida con menos divisiones. Este modelo debería compararse con un modelo que se entrenó desde el principio con menos divisiones. Si esto funciona de la forma esperada podemos proponer que el entrenamiento se haga en dos etapas, la primera con redundancia para la formación de características locales, y la segunda (sin redundancia) para una inferencia rápida y eficiente.

En las siguientes 4 figuras veremos una inferencia la izquierda y su máscara de segmentación correspondiente a la derecha. La máscara de segmentación se pone como una guía visual para apreciar lo bueno o malo de la inferencia (localización).

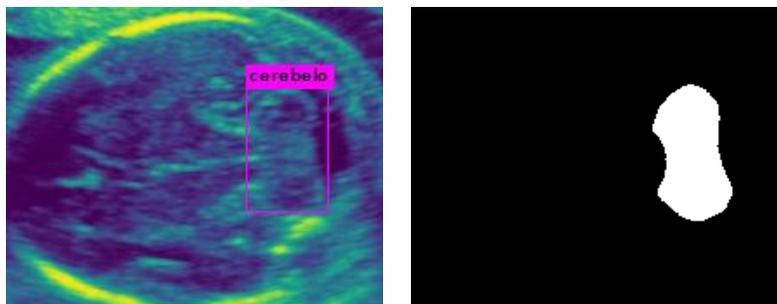


Figura 25: Inferencia con modelo de 49 divisiones, y después de 500 lotes, a la izquierda. Anotación experta a la derecha.

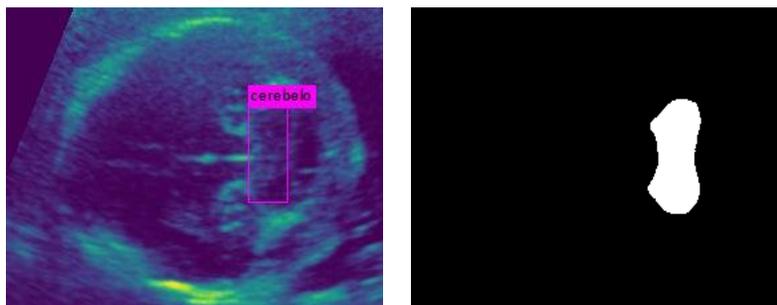


Figura 26: Inferencia con modelo de 81 divisiones, y después de 600 lotes, a la izquierda. Anotación experta a la derecha.

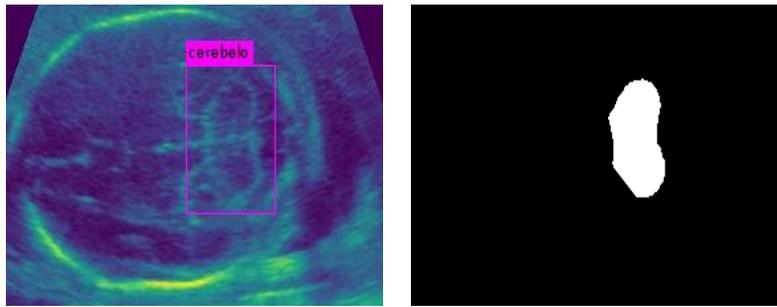


Figura 27: Inferencia con modelo de 121 divisiones, y después de 600 lotes, a la izquierda. Anotación experta a la derecha.

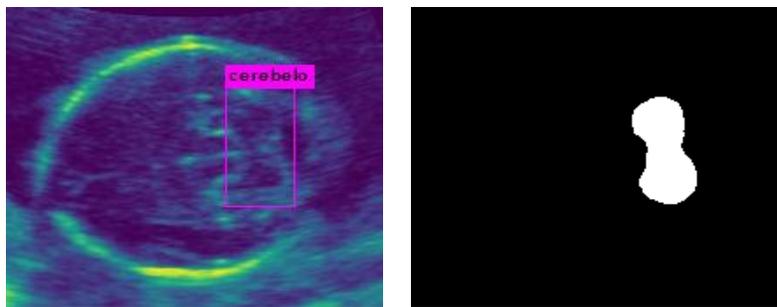


Figura 28: Inferencia con modelo de 169 divisiones, y después de 1000 lotes, a la izquierda. Anotación experta a la derecha.

Tiny YOLO

Se programaron 8 entrenamientos y se realizaron las inferencias con los datos de prueba para evaluar detección y localización. Se probaron 4 valores para el coeficiente de momento μ y 2 valores para la tasa de aprendizaje ϵ . La evaluación se hizo también calculando precisión promedio y puntaje F1 para las imágenes de prueba. Se reportan en la tabla enseguida los mejores resultados obtenidos en cada experimento (columna).

Momento μ	0.3		0.5		0.7		0.9	
Tasa de aprendizaje ϵ	$5 \cdot 10^{-4}$	$5 \cdot 10^{-3}$						
Precisión promedio	0.45	0.48	0.42	0.31	0.45	0.29	0.40	0.66
Puntaje F1	0.96	0.96	0.99	0.99	0.94	0.94	1	1

Tabla 2: Mejores resultados con Tiny YOLO (8 capas) para ocho entrenamientos.

La observación de la tabla 2 con los mejores resultados obtenidos nos dice que Tiny YOLO perdió la capacidad para localizar el cerebello con precisión, pero tuvo buenos resultados en detección.

La evolución de las métricas (precisión promedio y puntaje F1) mostraron un comportamiento caótico y por esto no se mostraron. Este comportamiento es típico de un entrenamiento con una tasa de aprendizaje demasiado grande. Ahora tenemos que preguntarnos porqué la tasa empleada en YOLO resultó demasiado grande para Tiny YOLO. Hay que investigar si la causa puede encontrarse en la reducción de la arquitectura, en la reducción del tamaño de entrada, o en ambas.

En la siguiente figura se muestran dos gráficas para dos tasas de aprendizaje. En cada gráfica se puede ver la evolución de los falsos positivos para 4 valores de momento.

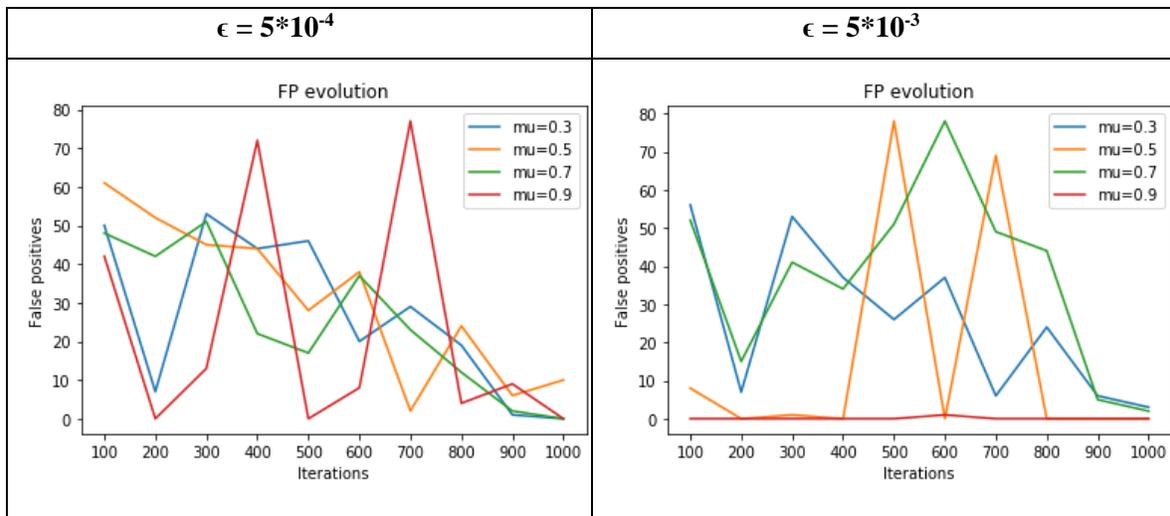


Figura 29: Evolución de falsos positivos para dos tasas de aprendizaje.

Se graficó la evolución de los falsos positivos donde se vio, en general, una tendencia a la baja o un comportamiento de aprendizaje. Se puede observar también en estas gráficas que la tasa de aprendizaje más grande permitió obtener la mayoría de los modelos con un mínimo de falsos positivos. Y una cosa más que debe llamar nuestra atención es el aprendizaje casi inmediato en algunos modelos con apenas una época de entrenamiento. Con estos experimentos no se pudo comprobar la aceleración del aprendizaje por lo inestable del mismo, pero si se comprobó la importancia de elegir una tasa de aprendizaje adecuada.

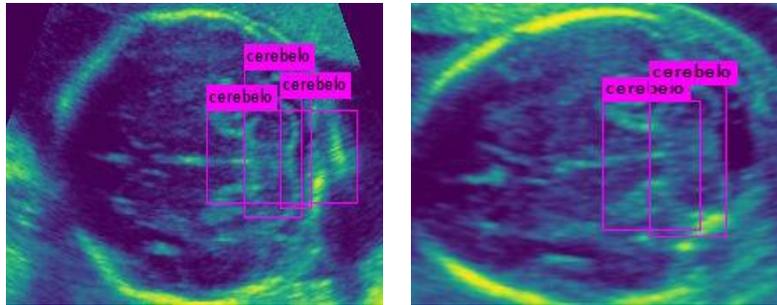


Figura 30: Ejemplos de inferencias con falsos positivos.

6.2 Segmentación semántica

Datos centrados y no centrados

Se entrenan 32 configuraciones guardando el mejor modelo para cada una. Luego se evalúan los 32 modelos guardados con precisión promedio (de la intersección sobre la unión). Las configuraciones combinan dos conjuntos de datos, dos arquitecturas CNN para extracción de características, 4 formas de actualizar los parámetros, y dos formas de aprendizaje. A continuación, una tabla con los 10 mejores resultados obtenidos al evaluar con datos de prueba.

conjunto	vgg	optimizador	aprendizaje	AP	épocas
centrado	16	Adadelta	transferencia	0.947	900
centrado	19	Adadelta	transferencia	0.947	500
centrado	19	Adadelta	desde cero	0.944	500
centrado	16	Adadelta	desde cero	0.94	450
centrado	19	Adam	desde cero	0.935	180
centrado	16	SGD	transferencia	0.932	900
centrado	19	SGD	transferencia	0.929	900
centrado	16	SGD	desde cero	0.921	900
centrado	16	Nadam	desde cero	0.916	600
centrado	16	Adam	desde cero	0.864	300

Tabla 3: Mejores 10 resultados luego de 32 experimentos.

Los resultados en la tabla 3 hacen evidente que FCN aprendió mejor la tarea de segmentar el cerebelo con los datos centrados y con el optimizador Adadelta. Entonces se dejaron de lado otros optimizadores y los datos originales para continuar con los experimentos. Los mejores resultados se obtuvieron con transferencia de aprendizaje, pero la diferencia es mínima. En estos mejores resultados las dos versiones de VGG están empatadas y, de forma inesperada, VGG19 (con más elementos) requirió menos entrenamiento.

Adadelata se caracteriza por no requerir un valor para la tasa de aprendizaje, pero si tener tasa de aprendizaje para cada dimensión. ¿Es esto particularmente útil para FCN? ¿Cómo?

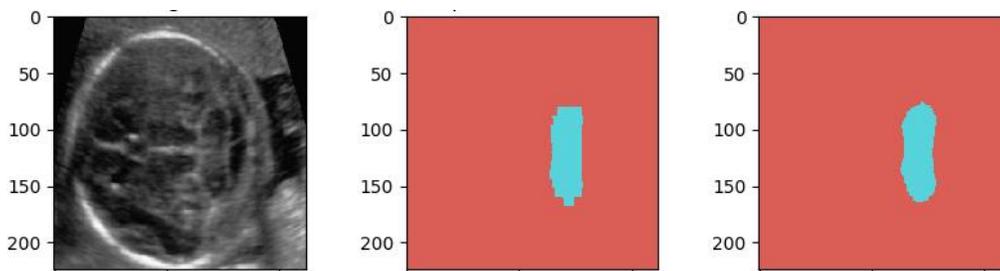


Figura 31: Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con datos no centrados, VGG19, optimizador Adadelata, y “desde cero”. El conjunto de pesos fue entrenado 400 épocas y obtuvo una evaluación AP=0.861.

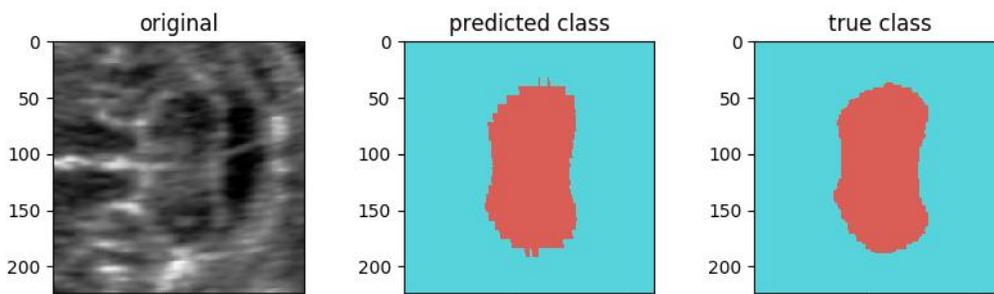


Figura 32: Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con datos centrados, VGG16, optimizador Adadelata, y con transferencia de aprendizaje. El conjunto de pesos fue entrenado 900 épocas y obtuvo una evaluación AP=0.947

Luego de experimentar con YOLO se decidió poner especial atención en el modo de entrenamiento o método de inicialización. Calculamos el promedio de la evaluación AP y del número de épocas para todos los modelos, de la primera serie, entrenados con transferencia y desde cero. Con estos datos podemos ver que el entrenamiento con transferencia es, por poco, mejor y más rápido.

Método	AP promedio	Épocas promedio
Transferencia	0.6683	419
Desde cero	0.6531	436

Tabla 4: Promedios para 2 métodos de inicialización de pesos.

Calculamos AP promedio y épocas promedio para experimentos realizados con VGG16 y VGG19. La elección del codificador significa tomar un número de capas convolucionales para

extracción de características. Observamos que un número mayor de capas puede disminuir el desempeño.

Codificador	AP promedio	Épocas promedio
VGG16	0.7039	440
VGG19	0.6174	410

Tabla 5: Promedios para dos configuraciones de CNN.

Datos aumentados por simetría

Se evalúan 8 modelos entrenados en la segunda serie de experimentos o sesiones de aprendizaje con FCN. El aprendizaje se apoya en recortes de medio cerebelo. Se prueban nuevamente las arquitecturas VGG16 y VGG19, y dos formas de aprendizaje. Además, se prueban dos formas de iniciar los valores cuando no son transferidos de un modelo aprendido. En la siguiente tabla se anotan los resultados.

vgg	inicialización	aprendizaje	AP	épocas
16	Glorot	transferencia	0.918	175
16	He	transferencia	0.918	170
19	He	desde cero	0.916	200
19	He	transferencia	0.916	160
19	Glorot	desde cero	0.915	240
16	Glorot	desde cero	0.914	200
19	Glorot	transferencia	0.914	160
16	He	desde cero	0.912	200

Tabla 6: Resultados con datos centrados y aumentados por simetría.

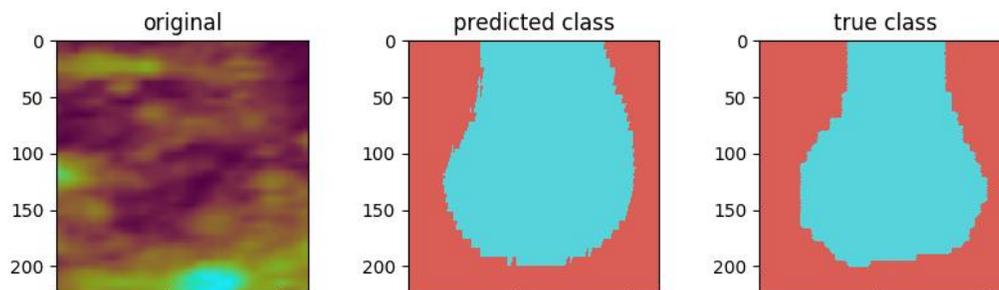


Figura 33: Imagen de entrada, segmentación automática, y anotación experta para el mejor modelo obtenido por inicialización de Glorot. AP=0.918

Los mejores resultados para la segunda serie de ensayos (tabla 6) muestran nuevamente que el entrenamiento con transferencia es marginalmente superior. Un objetivo particular para esta serie es de comparar dos formas de inicializar los elementos sin transferencia de aprendizaje. Observamos que en general se obtiene una precisión promedio más alta con He (3 posiciones entre las 4 mejores), sin embargo, Glorot y He están empatados en los mejores modelos obtenidos.

Datos clasificados

Antes que se calcularan los promedios donde se vio la superioridad de VGG16 frente a VGG19 se vio en las tablas 3 y 6 que VGG19 aparece 3 veces entre los primeros 5 lugares. Con este criterio decidimos dejar de lado VGG16 y continuar con los ensayos utilizando VGG19.

Finalmente se realizó una tercera serie de ajustes al modelo FCN-8 utilizando datos con un cuarto de cerebelo y disminuyendo el tamaño de entrada a 160*160. Los datos de entrada y las etiquetas corresponden a recortes de un cuarto de cerebelo. Se sigue investigando la forma de aprendizaje.

modo	lado	aprendizaje	AP	épocas
distal	160	desde cero	0.914	200
distal	160	transferencia	0.913	175
proximal	160	transferencia	0.855	175
proximal	160	desde cero	0.854	200

Tabla 7: Evaluación de aprendizaje con dos subconjuntos y cada uno con 18 subclases.

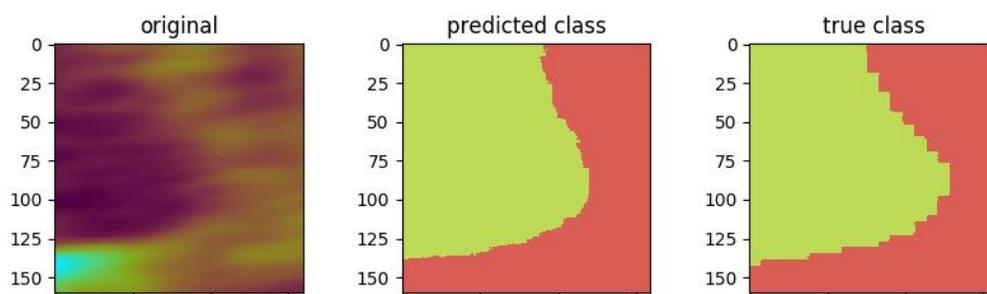


Figura 34: Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con cuartos distales y aprendizaje desde cero. El conjunto de pesos fue optimizado 200 épocas mediante Adadelta y obtuvo una evaluación AP=0.914.

En la última serie, con imágenes de cuarto de cerebelo, se observa una disminución del desempeño a comparación de la serie anterior, con datos de medio cerebelo. Se hace evidente que el modelo aprende mejor a segmentar objetos completos y no por partes.

En una serie extra de entrenamientos se repite la última serie, pero se regresa el tamaño de entrada a 224 x 224 píxeles.

datos	lado	aprendizaje	AP	épocas
distal	224	desde cero	0.896	200
distal	224	transferencia	0.895	190
proximal	224	desde cero	0.831	190
proximal	224	transferencia	0.82	175

Tabla 8: Evaluación de aprendizaje con dos subconjuntos de datos.

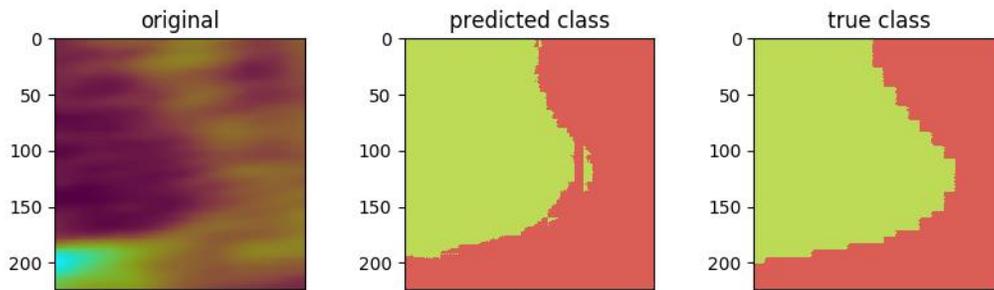


Figura 35: Imagen de entrada, segmentación de salida, y anotación experta para el modelo entrenado con datos distales. El conjunto de pesos fue entrenado 200 épocas desde cero y obtuvo una evaluación AP=0.896.

Con la serie extra caemos en cuenta que faltó poner más atención en el tamaño de entrada desde el principio. La escala del objeto y de sus características podría tener una afectación en el desempeño más importante que el modo de entrenamiento o la forma de iniciar los valores en general.

7. Conclusiones

Se implementó detección, localización y segmentación de cerebelo en imágenes de US fetal mediante aprendizaje profundo. Se realizaron diversos experimentos de aprendizaje automático a partir de imágenes de US fetal y anotaciones de expertos. Se transformaron las etiquetas y los datos buscando el mejor desempeño. Se investigó el aprendizaje con transferencia de conocimiento y el aprendizaje desde cero. Se experimentó con el número de capas en redes convolucionales encontrando que se puede tener un desempeño reducido por exceso de capas y que una falta de estas puede significar una pérdida importante en la funcionalidad. También se encontró que FCN obtiene la mejor precisión en imágenes donde el objeto tiene un lugar central y preponderante. Se ensayó la segmentación de cerebelo por partes, pero encontramos que las configuraciones de FCN probadas tienen mejores resultados segmentando objetos completos. La experiencia obtenida con este trabajo y la revisión del estado del arte nos conducen a proponer una variedad de trabajos a futuro.

Detección y localización

Se realizó la detección y localización automática de cerebelo en imágenes de US fetal aplicando aprendizaje de máquina en un modelo conexionista profundo. Investigamos el hiperparámetro S y encontramos que más grande es este número, mejor es la precisión en localización.

Se implementó transferencia de aprendizaje para tratar un conjunto de datos reducido (21 pacientes) y como una práctica común en aprendizaje profundo para imágenes médicas. Nos preguntamos sobre el efecto y la importancia de esta práctica por lo que nos propusimos experimentar con aprendizaje sin transferencia de pesos o aleatorios.

Se verificó la importancia del número de capas en una arquitectura CNN (24 en *YOLO*) al observar la pérdida de funcionalidad (localización) en una red *Tiny YOLO* con 8 capas. Proponemos investigar el número mínimo necesario de capas CNN (entre 9 y 23) para conservar resultados aceptables en localización.

Segmentación semántica

Se realizaron 32 experimentos con FCN-8 probando dos conjuntos de datos, dos variantes de codificador, dos modos de entrenamiento, y cuatro optimizadores diferentes. En estas pruebas se vio que los datos centrados y el optimizador Adadelta permiten obtener la mejor precisión en la segmentación. Esto significa que sí será conveniente utilizar un método de localización como YOLO para luego realizar la segmentación con FCN.

Luego se ejecutaron ocho entrenamientos adicionales probando datos aumentados por simetría. La intención era que la red aprendiera a segmentar medio cerebelo. Finalmente se probó la segmentación de diferentes grupos o modelos estadísticos de cuarto de cerebelo proximal y distal. Estas ideas no fracasaron del todo, pero tampoco permitieron mejorar los resultados. A partir

de los resultados se puede decir que FCN funciona mejor cuando aprende a segmentar objetos completos y en su contexto. Surgen algunas preguntas: ¿Porqué, en general, se obtuvieron mejores modelos con datos distales? ¿La idea de incluir un análisis en el aprendizaje es totalmente inadecuada o pudieron presentarse problemas en la implementación probada? ¿Y si cambamos el problema de segmentar el área por el de segmentar los puntos del contorno? ¿Sería de ayuda optimizar el número de capas para cada parte del análisis?

La experiencia obtenida en el presente trabajo indica que la transferencia de aprendizaje puede agilizar el entrenamiento y mejorar los resultados aun cuando los pesos se hayan entrenado con datos de otras clases sin relación aparente. Sin embargo, encontramos hiperparámetros que pueden tener una mayor importancia para el desempeño del modelo como puede ser el tamaño de entrada de la red y su relación con el tamaño de las imágenes a procesar.

En el caso del aprendizaje desde cero o totalmente aleatorio se comprobó que el método de inicialización de He es más apropiado que Glorot para trabajar con redes convolucionales profundas, en particular con el modelo de VGG.

Se probaron dos arquitecturas de codificador CNN pertenecientes a VGG16 y VGG19 y se observó que VGG16 tuvo, en promedio, un desempeño 14% superior. Dichos modelos CNN tienen 13 y 16 capas respectivamente. Se propone buscar el número de capas inferior a 13 y mínimo necesario para conservar una segmentación que permita una localización funcional.

Trabajo a futuro

El número de celdas S en YOLO está directamente relacionado a la capa densa y se puede intuir un posible efecto en el aprendizaje de las capas convolucionales. Se propone un aprendizaje en dos etapas para obtener modelos más eficientes en la inferencia.

Buscar las variantes de YOLO y FCN más eficientes y/o más precisas optimizando los hiperparámetros como un conjunto de características. Probar otras arquitecturas y proponer híbridos. Proponer un diseño con hiperparámetros obtenidos a partir de intuiciones humanas. Experimentar con modelos concatenados como se verá a continuación.

De forma análoga al modelo apilado de FCN visto en la figura 3.3 proponemos un modelo apilado YOLO-FCN para segmentación de cerebelo fetal. Se comprobó que se obtiene una segmentación más precisa al utilizar datos centrados. Estos datos se podrán obtener al aplicar YOLO y predecir la localización del cerebelo. Las predicciones de YOLO incluyen las coordenadas y las dimensiones. Con esta información podemos recortar el cerebelo (más un margen) y pasarlo a FCN para obtener la segmentación.

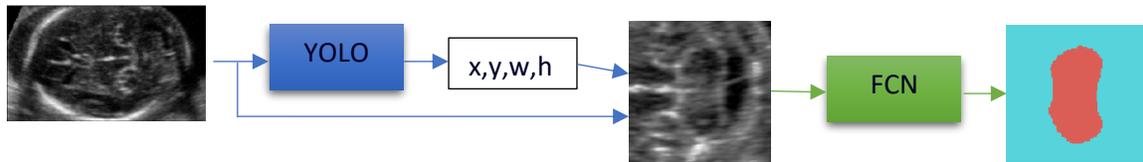


Figura 36: Propuesta de modelo apilado YOLO-FCN para segmentación fina del cerebro.

Con la precisión obtenida en la segmentación y la información de la escala en la imagen de ultrasonido parece factible proponer un sistema automatizado para la biometría del cerebro y la estimación de la edad gestacional. A continuación, un esquema para dicho sistema.

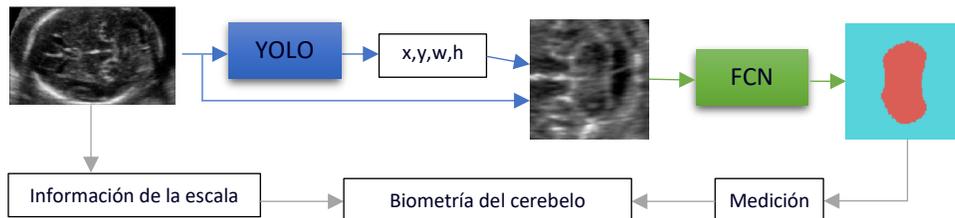


Figura 37: Propuesta de sistema automatizado para la biometría de cerebro.

También se podría utilizar la localización de YOLO para un método tradicional de aprendizaje de máquina como ASM que requiere una inicialización para su funcionamiento correcto.

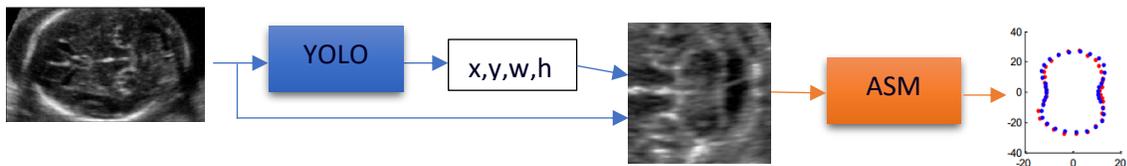


Figura 38: Propuesta de modelo apilado YOLO-ASM.

En el estado del arte se vio como Sundaresan et al. utilizaron FCN para clasificación de planos estándar. Considerando este trabajo y nuestra propuesta para encontrar una red tipo VGG con un número mínimo de capas podemos anticipar un diseño donde FCN se utiliza para localización.

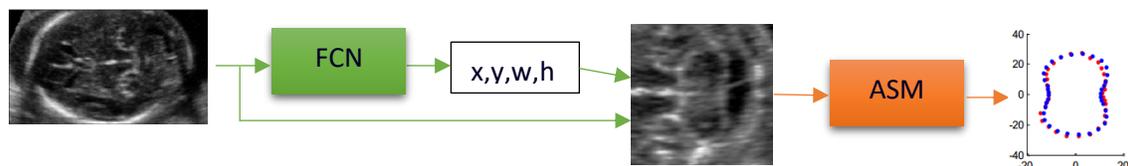


Figura 39: Propuesta de modelo apilado FCN-ASM.

Finalmente habrá que ensayar con cambios de paradigma o funcionalidad. Por ejemplo, se pueden buscar los puntos del contorno como única clase o como tercera clase en una segmentación o por sus coordenadas en una predicción. Dada la importancia que se encontró en la relación de dimensiones de entrada y de datos, también será deseable incluir un análisis multi escala y/o probar una entrada por ventana deslizante.

Bibliografía

- [1] R. Houston, S. Parsons, G. Jones, y A. Benett, «Biosonar: Seeing with Sound», mar. 06, 2001. <http://www.biosonar.bris.ac.uk/index.htm> (accedido jun. 29, 2020).
- [2] J. Woo, «History of Ultrasound in Obstetrics and Gynecology, Part 1», nov. 2006. <http://www.ob-ultrasound.net/history1.html> (accedido jun. 29, 2020).
- [3] J. Woo, «History of Ultrasound in Obstetrics and Gynecology, Part 2», nov. 2006. <http://www.ob-ultrasound.net/history2.html> (accedido jun. 29, 2020).
- [4] J. Woo, «History of Ultrasound in Obstetrics and Gynecology, Part 3», nov. 2006. <http://www.ob-ultrasound.net/history3.html> (accedido jun. 29, 2020).
- [5] «Sonographic examination of the fetal central nervous system: guidelines for performing the 'basic examination' and the 'fetal neurosonogram'», *Ultrasound Obstet Gynecol*, vol. 29, n.º 1, pp. 109-116, ene. 2007, doi: 10.1002/uog.3909.
- [6] A. Sepúlveda-Martínez y V. Sepúlveda, «Biometría de cerebelo fetal: ¿Parámetro útil en edad gestacional dudosa?», *Revista chilena de Ultrasonografía*, vol. 10, pp. 122-127, ene. 2007.
- [7] V. Bunduki, M. Zugaib, y V. Bunduki, *Atlas of fetal ultrasound: normal imaging and malformations*. 2018.
- [8] X. Güell Paradís, «The role of the cerebellum in movement, thought, and emotion», Ph.D. Thesis, Universitat Autònoma de Barcelona, 2018.
- [9] S. Skansi, *Introduction to deep learning: from logical calculus to artificial intelligence*. Cham, Switzerland: Springer, 2018.
- [10] V. Peckhaus, «Calculus Ratiocinator versus Characteristica Universalis? The Two Traditions in Logic, Revisited», *History and Philosophy of Logic*, vol. 25, n.º 1, pp. 3-14, feb. 2004.
- [11] J. S. Mill, «A System Of Logic, Ratiocinative And Inductive», p. 1182.
- [12] J. McCarthy, M. L. Minsky, N. Rochester, y C. E. Shannon, *A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE*. 1955.
- [13] B. D. R. Needham, «Lighthill Report: Artificial Intelligence: a paper symposium».
- [14] Alberto Greco, «Cognitive science and cognitive sciences», *Journal of Cognitive Science*, vol. 13, n.º 4, pp. 471-485, dic. 2012, doi: 10.17791/JCS.2012.13.4.471.
- [15] B. Macukow, «Neural Networks – State of Art, Brief History, Basic Models and Architecture», en *Computer Information Systems and Industrial Management*, vol. 9842, K. Saeed y W. Homenda, Eds. Cham: Springer International Publishing, 2016, pp. 3-14.
- [16] W. S. McCulloch y Pitts, *W.: A logical calculus of the ideas immanent in nervous activity*. .
- [17] R. C. Tees, «Review of The organization of behavior: A neuropsychological theory.», *Canadian Psychology/Psychologie canadienne*, vol. 44, n.º 1, pp. 74-76, 2003, doi: 10.1037/h0088061.
- [18] E. Stellar, *The Quarterly Review of Biology*, vol. 26, n.º 3, pp. 319-320, 1951.
- [19] I. T. Brain y F. Rosenblatt, *The Perceptron: A Probabilistic Model for Information Storage and Organization*. .
- [20] F. Rosenblatt, *Principles of Neurodynamics*. Washington: Spartan Books, 1962.
- [21] K. Fukushima, «Cognitron: a self-organising multilayered neural network», *Biological Cybernetics*, vol. 20, n.º 3, pp. 127-136, ene. 1975.
- [22] S. Grossberg, «Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors.», *Biological cybernetics*, vol. 23, n.º 3, pp. 121-134, jul. 1976.
- [23] R. Heath, «Parallel models of associative memory: G. Hinton and J. Anderson (Eds.). Hillsdale, N. J. Erlbaum, 1981. pp. viii + 295. \$26.20.», *Journal of Mathematical Psychology*, vol. 26, n.º 1, pp. 78-88, 1982.

- [24] T. Kohonen, «Self-organized formation of topologically correct feature maps», *Biol. Cybern.*, vol. 43, n.º 1, pp. 59-69, 1982, doi: 10.1007/BF00337288.
- [25] «Biblioteca Digital UNAM - Inicio». <https://www.bidi.unam.mx/> (accedido jun. 17, 2020).
- [26] H. A. Babri y Yin Tong, «Deep feedforward networks: application to pattern recognition», *Proceedings of International Conference on Neural Networks (ICNN'96), Neural Networks, 1996., IEEE International Conference on*, vol. 3, p. 1422, ene. 1996, doi: 10.1109/ICNN.1996.549108.
- [27] S. Hochreiter y J. Schmidhuber, «Long Short-Term Memory», *Neural Computation*, vol. 9, n.º 8, pp. 1735-1780, nov. 1997, doi: 10.1162/neco.1997.9.8.1735.
- [28] E. Çevik y K. Zengin, «Classification of Skin Lesions in Dermatoscopic Images with Deep Convolution Network», *European Journal of Science and Technology*, pp. 309-318, 2019, doi: 10.31590/ejosat.638247.
- [29] M. Roby y C. Jayaraman, «Segmentation technique for medical image processing: A survey», 2017, pp. 1055-1061, doi: 10.1109/ICICI.2017.8365301.
- [30] I. Aizenberg, N. Aizenberg, C. Butakoff, y E. Farberov, «Image Recognition on the Neural Network Based on Multi-Valued Neurons.», 2000, vol. 2, pp. 2989-2992, doi: 10.1109/ICPR.2000.906241.
- [31] P. Picton, «Threshold Logic», en *Introduction to Neural Networks*, London: Macmillan Education UK, 1994, pp. 133-142.
- [32] S. L. Hurst, «An introduction to threshold logic: a survey of present theory and practice», *Radio and Electronic Engineer*, vol. 37, n.º 6, pp. 339-351(12), jun. 1969.
- [33] J. Laserson, «From Neural Networks to Deep Learning: zeroing in on the human brain», *XRDS*, vol. 18, n.º 1, pp. 29-34, sep. 2011, doi: 10.1145/2000775.2000787.
- [34] «Geoffrey Hinton: Talks available on the web». <https://www.cs.toronto.edu/~hinton/talks.html> (accedido jun. 25, 2020).
- [35] G. E. Hinton, «Reducing the Dimensionality of Data with Neural Networks», *Science*, vol. 313, n.º 5786, pp. 504-507, jul. 2006, doi: 10.1126/science.1127647.
- [36] Y. Bengio, P. Lamblin, D. Popovici, y H. Larochelle, «Greedy Layer-Wise Training of Deep Networks», en *Advances in Neural Information Processing Systems 19*, B. Schölkopf, J. C. Platt, y T. Hoffman, Eds. MIT Press, 2007, pp. 153–160.
- [37] S. Liu *et al.*, «Deep Learning in Medical Ultrasound Analysis: A Review», *Engineering*, vol. 5, n.º 2, pp. 261-275, abr. 2019, doi: 10.1016/j.eng.2018.11.020.
- [38] Yuya Hiramatsu, Chisako Muramatsu, Hironobu Kobayashi, Takeshi Hara, y Hiroshi Fujita, «Automated detection of masses on whole breast volume ultrasound scanner: false positive reduction using deep convolutional neural network», mar. 2017, vol. 10134, doi: 10.1117/12.2254581.
- [39] C. Bian, R. Lee, Y.-H. Chou, y J.-Z. Cheng, «Boundary Regularized Convolutional Neural Network for Layer Parsing of Breast Anatomy in Automated Whole Breast Ultrasound», en *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, Cham, 2017, pp. 259-266.
- [40] J. Shi, S. Zhou, X. Liu, Q. Zhang, M. Lu, y T. Wang, «Stacked deep polynomial network based representation learning for tumor classification with small ultrasound image dataset», *Neurocomputing*, vol. 194, pp. 87-94, jun. 2016, doi: 10.1016/j.neucom.2016.01.074.
- [41] S. Azizi *et al.*, «Ultrasound-Based Detection of Prostate Cancer Using Automatic Feature Selection with Deep Belief Networks», en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, Cham, 2015, pp. 70-77.
- [42] X. Yang *et al.*, *Fine-grained Recurrent Neural Networks for Automatic Prostate Segmentation in Ultrasound Images*. 2016.

- [43] K. Wu, X. Chen, y M. Ding, «Deep learning based classification of focal liver lesions with contrast-enhanced ultrasound», *Optik*, vol. 125, n.º 15, pp. 4057-4063, ago. 2014, doi: 10.1016/j.ijleo.2014.01.114.
- [44] F. C. Ghesu, B. Georgescu, Y. Zheng, J. Hornegger, y D. Comaniciu, «Marginal Space Deep Learning: Efficient Architecture for Detection in Volumetric Image Data», en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, Cham, 2015, pp. 710-718.
- [45] Franklin Pereira *et al.*, «Automated detection of coarctation of aorta in neonates from two-dimensional echocardiograms», *Journal of Medical Imaging*, vol. 4, n.º 1, pp. 1-13, ene. 2017, doi: 10.1117/1.JMI.4.1.014502.
- [46] P. Sombune, P. Phienphanich, S. Phuechpanpaisal, S. Muengtawepongsa, A. Ruamthanthong, y C. Tantibundhit, «Automated embolic signal detection using Deep Convolutional Neural Network.», *Conf Proc IEEE Eng Med Biol Soc*, vol. 2017, pp. 3365-3368, jul. 2017, doi: 10.1109/EMBC.2017.8037577.
- [47] F. Milletari *et al.*, «Robust Segmentation of Various Anatomies in 3D Ultrasound Using Hough Forests and Learned Data Representations», en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015*, Cham, 2015, pp. 111-118.
- [48] K. Lekadir *et al.*, «A Convolutional Neural Network for Automatic Characterization of Plaque Composition in Carotid Ultrasound.», *IEEE J Biomed Health Inform*, vol. 21, n.º 1, pp. 48-55, ene. 2017, doi: 10.1109/JBHI.2016.2631401.
- [49] J. Y. Shin, N. Tajbakhsh, R. T. Hurst, C. B. Kendall, y J. Liang, «Automating Carotid Intima-Media Thickness Video Interpretation with Convolutional Neural Networks», en *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, jun. 2016, pp. 2526-2535, doi: 10.1109/CVPR.2016.277.
- [50] J. Ma, F. Wu, T. Jiang, J. Zhu, y D. Kong, «Cascade convolutional neural networks for automatic detection of thyroid nodules in ultrasound images», *Medical Physics*, vol. 44, n.º 5, pp. 1678-1691, may 2017, doi: 10.1002/mp.12134.
- [51] E. Smistad y L. Løvstakken, «Vessel Detection in Ultrasound Images Using Deep Convolutional Neural Networks», en *Deep Learning and Data Labeling for Medical Applications*, Cham, 2016, pp. 30-38.
- [52] M. Yaqub, B. Kelly, A. T. Papageorghiou, y J. A. Noble, «A Deep Learning Solution for Automatic Fetal Neurosonographic Diagnostic Plane Verification Using Clinical Standard Constraints», *Ultrasound in Medicine & Biology*, vol. 43, n.º 12, pp. 2925-2933, dic. 2017, doi: 10.1016/j.ultrasmedbio.2017.07.013.
- [53] W. Huang, C. P. Bridge, J. A. Noble, y A. Zisserman, «Temporal HeartNet: Towards Human-Level Automatic Analysis of Fetal Cardiac Screening Video», en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2017*, Cham, 2017, pp. 341-349.
- [54] Y. Gao y J. Alison Noble, «Detection and Characterization of the Fetal Heartbeat in Free-hand Ultrasound Sweeps with Weakly-supervised Two-streams Convolutional Networks», en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2017*, Cham, 2017, pp. 305-313.
- [55] H. Qi, S. Collins, y A. Noble, «Weakly Supervised Learning of Placental Ultrasound Images with Residual Networks», en *Medical Image Understanding and Analysis*, Cham, 2017, pp. 98-108.
- [56] H. Chen, Y. Zheng, J.-H. Park, P.-A. Heng, y S. K. Zhou, «Iterative Multi-domain Regularized Deep Learning for Anatomical Structure Detection and Segmentation from Ultrasound Images», en *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2016*, Cham, 2016, pp. 487-495.
- [57] H. Ravishankar, S. M. Prabhu, V. Vaidya, y N. Singhal, «Hybrid approach for automatic segmentation of fetal abdomen from ultrasound images using deep learning», en *2016 IEEE*

- 13th International Symposium on Biomedical Imaging (ISBI), abr. 2016, pp. 779-782, doi: 10.1109/ISBI.2016.7493382.
- [58] Y. Zhang, M. T. C. Ying, L. Yang, A. T. Ahuja, y D. Z. Chen, «Coarse-to-Fine Stacked Fully Convolutional Nets for lymph node segmentation in ultrasound images», en *2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, dic. 2016, pp. 443-448, doi: 10.1109/BIBM.2016.7822557.
- [59] H. Ravishankar, R. Venkataramani, S. Thiruvankadam, P. Sudhakar, y V. Vaidya, «Learning and Incorporating Shape Models for Semantic Segmentation», en *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, Cham, 2017, pp. 203-211.
- [60] J. Hetherington, V. Lessoway, V. Gunka, P. Abolmaesumi, y R. Rohling, «SLIDE: automatic spine level identification system using a deep convolutional neural network», *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, n.º 7, pp. 1189-1198, jul. 2017, doi: 10.1007/s11548-017-1575-8.
- [61] D. Golan, Y. Donner, C. Mansi, J. Jaremko, y M. Ramachandran, «Fully Automating Graf's Method for DDH Diagnosis Using Deep Convolutional Neural Networks», en *Deep Learning and Data Labeling for Medical Applications*, Cham, 2016, pp. 130-141.
- [62] A. R. Hareendranathan *et al.*, «Toward automatic diagnosis of hip dysplasia from 2D ultrasound», en *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, abr. 2017, pp. 982-985, doi: 10.1109/ISBI.2017.7950680.
- [63] P. Burlina, S. Billings, N. Joshi, y J. Albayda, «Automated diagnosis of myositis from muscle ultrasound: Exploring the use of machine learning and deep learning methods», *PLOS ONE*, vol. 12, n.º 8, p. e0184059, ago. 2017, doi: 10.1371/journal.pone.0184059.
- [64] A. Hafiane, P. Vieyres, y A. Delbos, «Deep learning with spatiotemporal consistency for nerve segmentation in ultrasound images», jun. 2017.
- [65] Z. Cao *et al.*, «Breast Tumor Detection in Ultrasound Images Using Deep Learning», en *Patch-Based Techniques in Medical Imaging*, Cham, 2017, pp. 121-128.
- [66] C. F. Baumgartner, K. Kamnitsas, J. Matthew, S. Smith, B. Kainz, y D. Rueckert, «Real-Time Standard Scan Plane Detection and Localisation in Fetal Ultrasound Using Fully Convolutional Neural Networks», en *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016*, Cham, 2016, pp. 203-211.
- [67] F. T. Dezaki *et al.*, «Deep Residual Recurrent Neural Networks for Characterisation of Cardiac Cycle Phase from Echocardiograms», en *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Cham, 2017, pp. 100-108.
- [68] Andrew R. Jamieson, Karen Drukker, y Maryellen L. Giger, «Breast image feature learning with adaptive deconvolutional networks», feb. 2012, vol. 8315, doi: 10.1117/12.910710.
- [69] J.-Z. Cheng *et al.*, «Computer-Aided Diagnosis with Deep Learning Architecture: Applications to Breast Lesions in US Images and Pulmonary Nodules in CT Scans», *Scientific Reports*, vol. 6, n.º 1, p. 24454, abr. 2016, doi: 10.1038/srep24454.
- [70] S. Han *et al.*, «A deep learning framework for supporting the classification of breast lesions in ultrasound images», *Physics in Medicine & Biology*, vol. 62, n.º 19, pp. 7714-7728, sep. 2017, doi: 10.1088/1361-6560/aa82ec.
- [71] J. Ferlay, H.-R. Shin, F. Bray, D. Forman, C. Mathers, y D. M. Parkin, «Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008», *International Journal of Cancer*, vol. 127, n.º 12, pp. 2893-2917, dic. 2010, doi: 10.1002/ijc.25516.
- [72] J. Ma, F. Wu, J. Zhu, D. Xu, y D. Kong, «A pre-trained convolutional neural network based method for thyroid nodule diagnosis», *Ultrasonics*, vol. 73, pp. 221-230, ene. 2017, doi: 10.1016/j.ultras.2016.09.011.

- [73] T. Liu, S. Xie, J. Yu, L. Niu, y W. Sun, «Classification of thyroid nodules in ultrasound images using deep model based transfer learning and hybrid features», en *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, mar. 2017, pp. 919-923, doi: 10.1109/ICASSP.2017.7952290.
- [74] L. Wu, J. Cheng, S. Li, B. Lei, T. Wang, y D. Ni, «FUIQA: Fetal Ultrasound Image Quality Assessment With Deep Convolutional Networks», *IEEE Transactions on Cybernetics*, vol. 47, n.º 5, pp. 1336-1349, may 2017, doi: 10.1109/TCYB.2017.2671898.
- [75] V. Sundaresan, C. P. Bridge, C. Ioannou, y J. A. Noble, «Automated characterization of the fetal heart in ultrasound images using fully convolutional neural networks», en *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, abr. 2017, pp. 671-674, doi: 10.1109/ISBI.2017.7950609.
- [76] T. Acharya y A. K. Ray, *Image Processing - Principles and Applications*. USA: Wiley-Interscience, 2005.
- [77] L. Yu, Y. Guo, Y. Wang, J. Yu, y P. Chen, «Segmentation of Fetal Left Ventricle in Echocardiographic Sequences Based on Dynamic Convolutional Neural Networks.», *IEEE Trans Biomed Eng*, vol. 64, n.º 8, pp. 1886-1895, ago. 2017, doi: 10.1109/TBME.2016.2628401.
- [78] N. Baka, S. Leenstra, y T. van Walsum, «Ultrasound Aided Vertebral Level Localization for Lumbar Surgery», *IEEE Transactions on Medical Imaging*, vol. 36, n.º 10, pp. 2138-2147, 2017.
- [79] L. Wu, Y. Xin, S. Li, T. Wang, P. Heng, y D. Ni, «Cascaded Fully Convolutional Networks for automatic prenatal ultrasound image segmentation», en *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 663-666.
- [80] N. Singhal, S. Mukherjee, y C. Perrey, «Automated assessment of endometrium from transvaginal ultrasound using Deep Learned Snake», en *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, 2017, pp. 283-286.
- [81] M. Pesteie, V. Lessoway, P. Abolmaesumi, y R. N. Rohling, «Automatic Localization of the Needle Target for Ultrasound-Guided Epidural Injections», *IEEE Transactions on Medical Imaging*, vol. 37, n.º 1, pp. 81-92, 2018.
- [82] X. Yang *et al.*, «Towards Automatic Semantic Segmentation in Volumetric Ultrasound», en *Medical Image Computing and Computer Assisted Intervention – MICCAI 2017*, Cham, 2017, pp. 711–719.
- [83] V. A. B. Guerrero, «Desarrollo del estado del arte en investigación: una herramienta basada en inteligencia artificial», *Revista Politécnica*, vol. 15, n.º 30, pp. 70–81, 2019.
- [84] Md. Z. Alom *et al.*, «A State-of-the-Art Survey on Deep Learning Theory and Architectures», *Electronics*, vol. 8, p. 292, 2019, doi: 10.3390/electronics8030292.
- [85] A. Torfi, R. A. Shirvani, Y. Keneshloo, N. Tavaf, y E. A. Fox, «Natural Language Processing Advancements By Deep Learning: A Survey», *arXiv:2003.01200 [cs]*, jun. 2020, Accedido: sep. 28, 2020. [En línea]. Disponible en: <http://arxiv.org/abs/2003.01200>.
- [86] J. Chai y A. Li, «Deep Learning in Natural Language Processing: A State-of-the-Art Survey», en *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*, 2019, pp. 1-6.
- [87] «Application of deep learning to cybersecurity: A survey | Elsevier Enhanced Reader». <https://reader.elsevier.com/reader/sd/pii/S0925231219302954?token=1EECEBFFDF09E2342C011748C66D7ECB16BFF3A3736E7B61A776FB494092C3C1C3AB9666FCD0FB9C1930C55B1E332EAB> (accedido oct. 08, 2020).
- [88] G. V. Rodríguez, F. A. Cosío, M. E. G. Huerta, L. C. Marín, H. B. Olivares, y B. E. Ramírez, «Segmentación automática del cerebelo en volúmenes de ultrasonido del cerebro fetal», *Rev Mex Ing Biomed*, vol. 36, n.º 2, pp. 121-129, may 2015.
- [89] N. Qian, *On the Momentum Term in Gradient Descent Learning Algorithms*. 1999.

- [90] J. Redmon, S. Divvala, R. Girshick, y A. Farhadi, «You Only Look Once: Unified, Real-Time Object Detection», *arXiv:1506.02640 [cs]*, may 2016, Accedido: ago. 06, 2020. [En línea]. Disponible en: <http://arxiv.org/abs/1506.02640>.
- [91] J. Yan, Z. Lei, L. Wen, y S. Li, «The Fastest Deformable Part Model for Object Detection», 2014, doi: 10.1109/CVPR.2014.320.
- [92] R. Girshick, J. Donahue, T. Darrell, y J. Malik, «Rich feature hierarchies for accurate object detection and semantic segmentation», *arXiv:1311.2524 [cs]*, oct. 2014, Accedido: ago. 07, 2020. [En línea]. Disponible en: <http://arxiv.org/abs/1311.2524>.
- [93] G. Van Rossum y F. L. Drake, *Python 3 Reference Manual*. Scotts Valley, CA: CreateSpace, 2009.
- [94] J. Redmon, *Darknet: Open Source Neural Networks in C*. 2013.
- [95] M. Fatica, «CUDA toolkit and libraries», en *2008 IEEE Hot Chips 20 Symposium (HCS)*, ago. 2008, pp. 1-22, doi: 10.1109/HOTCHIPS.2008.7476520.
- [96] M. Everingham, S. M. A. Eslami, L. Van Gool, C. K. I. Williams, J. Winn, y A. Zisserman, «The Pascal Visual Object Classes Challenge: A Retrospective», *International Journal of Computer Vision*, vol. 111, n.º 1, pp. 98-136, ene. 2015, doi: 10.1007/s11263-014-0733-5.
- [97] J. D. Hunter, «Matplotlib: A 2D graphics environment», *Computing in Science & Engineering*, vol. 9, n.º 3, pp. 90–95, 2007, doi: 10.1109/MCSE.2007.55.
- [98] *A Better Default Colormap for Matplotlib | SciPy 2015 | Nathaniel Smith and Stéfan van der Walt*. 2015.
- [99] S.-H. Tsang, «Review: YOLOv1 — You Only Look Once (Object Detection)», *Medium*, mar. 20, 2019. <https://towardsdatascience.com/yolov1-you-only-look-once-object-detection-e1f3ffec8a89> (accedido nov. 16, 2020).
- [100] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, y Li Fei-Fei, «ImageNet: A large-scale hierarchical image database», en *2009 IEEE Conference on Computer Vision and Pattern Recognition*, jun. 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.
- [101] «YOLO: Real-Time Object Detection». <https://pjreddie.com/darknet/yolov1/> (accedido oct. 29, 2020).
- [102] J. Long, E. Shelhamer, y T. Darrell, «Fully Convolutional Networks for Semantic Segmentation», 2015, pp. 3431-3440, Accedido: may 28, 2020. [En línea]. Disponible en: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Long_Fully_Convolutional_Networks_2015_CVPR_paper.html.
- [103] K. Simonyan y A. Zisserman, *Very Deep Convolutional Networks for Large-Scale Image Recognition*. 2014.
- [104] F. Chollet y others, *Keras*. 2015.
- [105] «Learn about Fully Convolutional Networks for semantic segmentation». <https://fairyonice.github.io/Learn-about-Fully-Convolutional-Networks-for-semantic-segmentation.html> (accedido ago. 27, 2020).
- [106] Martín Abadi *et al.*, *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015.
- [107] X. Glorot y Y. Bengio, «Understanding the difficulty of training deep feedforward neural networks», Chia Laguna Resort, Sardinia, Italy, may 2010, vol. 9, pp. 249–256, [En línea]. Disponible en: <http://proceedings.mlr.press/v9/glorot10a.html>.
- [108] K. He, X. Zhang, S. Ren, y J. Sun, *Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification*. 2015.
- [109] M. D. Zeiler, *ADADELTA: An Adaptive Learning Rate Method*. 2012.
- [110] D. P. Kingma y J. Ba, «Adam: A Method for Stochastic Optimization», *arXiv:1412.6980 [cs]*, ene. 2017, Accedido: ago. 31, 2020. [En línea]. Disponible en: <http://arxiv.org/abs/1412.6980>.
- [111] T. Dozat, «Incorporating nesterov momentum into adam», 2016.

