

**IMPLEMENTACIÓN DE REDES NEURONALES CONVOLUCIONALES PARA
LA CLASIFICACIÓN AUTOMÁTICA DE PLANCTON**

DENISSE ANDREA PÉREZ BECERRA

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
BUCARAMANGA
2016**

**IMPLEMENTACIÓN DE REDES NEURONALES CONVOLUCIONALES PARA
LA CLASIFICACIÓN AUTOMÁTICA DE PLANCTON**

DENISSE ANDREA PÉREZ BECERRA

**Trabajo de grado para optar al título de
Ingeniera de Sistemas**

**DIRECTOR
Ph.D RAUL RAMOS POLLÁN
Escuela de Ingeniería de Sistemas e Informática**

**UNIVERSIDAD INDUSTRIAL DE SANTANDER
ESCUELA DE INGENIERÍA DE SISTEMAS E INFORMÁTICA
FACULTAD DE INGENIERÍAS FÍSICO-MECÁNICAS
BUCARAMANGA
2016**

TABLA DE CONTENIDO

INTRODUCCIÓN	11
1. PLANTEAMIENTO DEL PROBLEMA.....	13
2. OBJETIVOS.....	14
2.1 OBJETIVO GENERAL	14
2.2 OBJETIVOS ESPECÍFICOS.....	14
3. MARCO TEÓRICO	15
3.1 REDES NEURONALES ARTIFICIALES.....	15
3.1.1 Modelo de neurona artificial.....	15
3.1.2 Arquitectura de las RNA.....	17
3.1.3 Aprendizaje de las RNA.....	18
3.2 REDES NEURONALES CONVOLUCIONALES	19
3.2.1 Características distintivas	20
3.2.2 Arquitectura.....	21
3.2.3 Métodos de Regularización.....	25
3.3 PLANCTON	28
3.3.1 Importancia del plancton.....	28
3.3.2 Proliferación de plancton.....	28
3.3.3 Clasificación.....	29
3.3.4 Otras clasificaciones	31
3.3.5 Distribución del plancton.....	34
4. ESTADO DEL ARTE.....	36
5. METODOLOGÍA	38

5.1 Etapa de Bosquejo del Diseño de la Solución	39
5.2 Etapa de Desarrollo del Prototipo	39
5.3 Etapa de Implementación y Pruebas	39
5.4 Etapa de Documentación.....	39
6. DESARROLLO DEL PROYECTO	40
6.1 Fase 1.....	40
6.1.1 Procesamiento de imágenes.....	40
6.2 FASE 2 DISEÑO Y DESARROLLO	43
6.2.1 Diseño de la red.....	43
6.2.2 Desarrollo del prototipo.....	44
6.2.3 Diseño de la red.....	44
6.2.4 Desarrollo del prototipo.....	45
6.2.5 Diseño de la red.....	45
6.2.6 Desarrollo del prototipo.....	46
6.2.7 Diseño de la red.....	46
6.2.8 Desarrollo del prototipo.....	46
6.2.9 Diseño de la red.....	47
6.2.10 Desarrollo del prototipo.....	47
6.3 FASE DE ENTRENAMIENTO.....	47
6.3.1 Primera variación	47
6.3.2 Segunda rotación.....	47
7. RESULTADOS OBTENIDOS.....	48
8. CONCLUSIONES	52
REFERENCIAS BIBLIOGRÁFICAS.....	54
BIBLIOGRAFÍA.....	55

LISTA DE FIGURAS

Figura 1. Modelo de Neurona Artificial.	Pág. 16
Figura 2. Funciones de Activación	Pág. 17
Figura 3. Arquitectura de RNA	Pág. 18
Figura 4. Capas de una CNN en un arreglo tridimensional	Pág. 20
Figura 5. Neuronas de una capa convolucional conectada a su campo receptivo.	Pág. 22
Figura 6. Arquitectura típica de una CNN	Pág. 25
Figura 7. Imagen satelital de una proliferación de plancton, color verde claro en la costa de Islandia. Cortesía del Observatorio Terrestre de la NASA.	Pág. 29
Figura 8. Cadena Trófica del Plancton	Pág. 30
Tabla 1. Clasificación taxonómica del plancton	Pág. 32
Figura 9. Vista externa de FlowCybot	Pág. 37
Figura 10. Metodología de trabajo, metodología ágil experimental.	Pág. 38
Figura 11. Ejemplo del primer procesado de plancton en una imagen del dataset.	Pág. 40
Figura 12. Ejemplo del segundo filtro aplicado, la transformada de Fourier.	Pág. 41
Figura 13. Ejemplo del tercer filtro aplicado, Logaritmo en base 10.	Pág. 41
Figura 14. Ejemplo del cuarto filtro aplicado, Sharpen.	Pág. 42
Figura 15. Ejemplo del quinto filtro aplicado, filtro Emboss.	Pág. 42
Figura 16. Ejemplo del quinto filtro aplicado, Filtro de convolución.	Pág. 43
Figura 17. Arquitectura del primer diseño de red realizado.	Pág. 44
Figura 18. Arquitectura del segundo diseño de red.	Pág. 45
Figura 19. Arquitectura del tercer diseño de red.	Pág. 45
Figura 20. Arquitectura del cuarto diseño de red.	Pág. 46
Figura 21. Arquitectura del quinto diseño de red.	Pág. 47

Tabla 2. Resultados obtenidos	Pág. 48
Figura 22. Porcentaje de acierto en la red con mejor desempeño a lo largo de las 30 iteraciones.	Pág. 49
Figura 23. Características extraídas en la primera capa de convolución.	Pág. 50
Figura 24. Media de las imágenes con el primer procesado realizado.	Pág. 51

RESUMEN

TÍTULO: IMPLEMENTACIÓN DE REDES NEURONALES CONVOLUCIONALES PARA LA CLASIFICACIÓN AUTOMÁTICA DE PLANCTON.*

AUTOR: DENISSE ANDREA PÉREZ BECERRA**

PALABRAS CLAVES: Redes neuronales convolucionales, Plancton, Procesamiento de imágenes.

DESCRIPCIÓN:

El Centro de Ciencia Marina Hatfield en la Universidad Estatal de Oregon, recolectó y clasificó manualmente miles de imágenes de las 121 clases de plancton. Estos datos fueron puestos a disposición de los usuarios gracias a una competencia en línea presentada por Kaggle en su sitio web. Dicha competencia busca un método de clasificación automático que ayude a la predicción de la salud del océano en cualquier momento, al clasificar las imágenes de plancton y permitiendo así que los científicos analicen las imágenes para evaluar la distribución y población de las especies. Usarán esta clasificación para estudiar también las cadenas de comida marina, la pesca y la conservación oceánica.

Los métodos de clasificación han avanzado grandemente en la última década gracias al estudio de las redes neuronales y sus formas más complejas como las redes neuronales convolucionales, que tienen un desempeño altamente superior comparado con el de los métodos tradicionales del aprendizaje de máquina.

Los datos proporcionados fueron pre procesados y entrenados en varias redes neuronales convolucionales que fueron creadas y modificadas dados los resultados obtenidos, hasta que fue alcanzado un buen porcentaje de clasificación. Estas redes también fueron entrenadas con los datos originales y con unas modificaciones al pre procesado para comparar los resultados. La mejor red tuvo un desempeño de 81.2032% y su diseño puede ser enviado a los científicos correspondientes para que clasifiquen nuevas imágenes capturadas en cualquier momento.

* Trabajo de grado.

** Facultad de Ingenierías Físico-Mecánicas. Escuela de Ingeniería de Sistemas e Informática. Director PhD. Raúl Ramos Pollán.

ABSTRACT

TITLE: IMPLEMENTATION OF CONVOLUTIONAL NEURAL NETWORKS FOR AUTOMATED PLANKTON CLASSIFICATION.*

AUTHOR: DENISSE ANDREA PÉREZ BECERRA**

KEY WORDS: Convolutional neural networks, Plankton, Image processing.

DESCRIPTION:

The Hatfield Marine Science Center at Oregon State University collected and manually classified thousands of images of the 121 classes of plankton. This data were made available thanks to an online competition on December 2014 presented by Kaggle on their website. The competition is looking for a faster automated classification method that helps predicting the health of the ocean at any given time by classifying plankton images. Classifying the plankton images and getting to know the amount of specimens per specie, allows the scientists analyze the images to assess species populations and distribution. The scientists will use this classification to study marine food webs, fisheries and ocean conservation.

Classification methods have had a great advance in the last decade with the study of neural networks and more complex forms of these such as the convolutional neural networks. CNN have a superior performance compared to the traditional machine learning methods.

The given data were preprocessed and trained in various convolutional neural networks that were created and modified given the results obtained until a good accuracy percentage was achieved in the classification. These networks were also trained with the original data and another processing to see the improvement of it. The best neural network has an 81.2032% accuracy and its design can be given to the scientists for them to classify new images they capture at any given time.

* Bachelor thesis.

** Faculty of physical Mechanical Engineering. Systems Engineering and Informatics Faculty. Director Phd Raúl Ramos Pollán.

INTRODUCCIÓN

El tamaño de los datos crece proporcionalmente a la capacidad computacional para manejarlos, requiriendo así un buen aprovechamiento de recursos y tiempo en las investigaciones. Con esto en mente, se hace evidente una ineficiencia en la explotación de recursos en los casos en que la comunidad continúa con métodos tradicionales por falta de conocimiento, generando un costo significativo de tiempo al obtener resultados eficientes.

Las redes convolucionales profundas han conseguido altos niveles de desempeño en tareas de reconocimiento de imágenes. Desde sus primeras aplicaciones 2006, este método ha ido ampliando con gran éxito su campo de aplicación y en la mayoría de competencias la diferencia entre los mejores métodos se reduce a la arquitectura o la forma de entrenar las redes convolucionales.

De forma general, los métodos basados en redes convolucionales profundas suelen mostrar un desempeño superior métodos más tradicionales entre el 5% y 20% dependiendo del problema abordado. Para el conjunto MNIST, una regularización de redes neuronales usando DropConnect logró un error final de tan solo 0.21% DeepFace logró un acierto de 97.35%, un 27% mayor que el estado del arte hasta ese momento para el conjunto de datos Labeled Faces in the Wild. Google ganó la competencia ImageNet en 2014 con un error de 6.66% usando redes profundas. La competencia de ImageNet trata de clasificar 1.2 millones de imágenes en unas 10.000 categorías distintas.

El presente proyecto nace como solución a una competencia en kaggle, desde el "National Data Science Bowl" que consiste en crear un algoritmo para automatizar el proceso de identificación, que será usado por científicos del Centro de Ciencia Marina de Hatfield, Oregon para estudiar las redes alimentarias marinas y conservación oceánica entre otros. El dataset disponible en Kaggle contiene 29860 imágenes de plancton en 121 categorías distintas.

Al probar varias combinaciones paramétricas y distinta cantidad y tipos de capas, fue posible llegar a un buen nivel de clasificación para la mayoría de tipos de plancton, a pesar que muchos de ellos contaban con pocos especímenes para la clasificación, mientras otros al ser bastante comunes presentaban gran cantidad de imágenes para entrenar, lo que representa un gran desbalanceo para la red al

momento del entrenamiento pues aprende con mayor rapidez, facilidad y robustez, las características cuando tuvo más ejemplos para aprender.

1. PLANTEAMIENTO DEL PROBLEMA

El plancton es de vital importancia para nuestro ecosistema, representando más de la mitad de la productividad primaria en la tierra y casi la mitad del total de carbono fijo en el ciclo global del carbono. Es la base de las redes de alimentación acuática, incluyendo las de las grandes pesquerías. La pérdida o disminución significativa de las poblaciones de plancton podría dar lugar a trastornos ecológicos, así como impactos sociales negativos, sobre todo en las culturas indígenas y las sociedades en desarrollo. La importancia global del plancton hace de sus niveles poblacionales una medida ideal de la salud de los océanos y ecosistemas a nivel mundial.

Los métodos tradicionales para medir y monitorear las poblaciones de plancton al ser realizados manualmente por análisis microscópico, consumen mucho tiempo y no pueden ser escaladas a la granularidad o alcance necesario para estudios a gran escala. Es necesario mejorar la forma de abordar dichas medidas, que hasta ahora se han realizado manualmente. Actualmente se están capturando las imágenes de éstos por medio de un sensor subacuático de imágenes, donde un sistema de cámaras realiza una captura de alta resolución a nivel microscópico sobre las grandes áreas de estudio, para ser analizadas posteriormente y determinar las poblaciones de especies y distribuciones de éstas.

El análisis manual de la cantidad de datos no es realmente factible. Tomaría alrededor de un año estudiar y clasificar el volumen de imágenes capturadas en un solo día, por tanto la clasificación automática de éstas puede ser automatizada por medio de métodos de aprendizaje de máquina que tendrían amplias aplicaciones para la evaluación de la salud oceanográfica y del ecosistema.

2. OBJETIVOS

2.1 OBJETIVO GENERAL

Evaluar el desempeño de las redes convolucionales profundas para la clasificación de imágenes de microscopía de alta resolución de plancton.

2.2 OBJETIVOS ESPECÍFICOS

- Implementar y evaluar distintas estrategias de pre procesamiento de imágenes previas al uso de las redes convolucionales.
- Implementar y evaluar distintas estructuras de redes convolucionales para la clasificación del plancton.
- Implementar y evaluar distintas estrategias de entrenamiento de redes convolucionales.
- Obtener métricas de los recursos computacionales necesarios para el entrenamiento y ejecución de los procesos implementados.
- Comparar el resultado obtenido con los resultados en la literatura y evaluar la conveniencia del uso de dichas redes en este problema.

3. MARCO TEÓRICO

3.1 REDES NEURONALES ARTIFICIALES

El aprendizaje automático es un campo de la ciencia de computación, que evolucionó del estudio de reconocimiento de patrones y la teoría computacional de aprendizaje de la inteligencia artificial. Explora el estudio y la construcción de algoritmos que puedan aprender y hacer predicciones de los datos. Dichos algoritmos operan construyendo un modelo con ejemplos de entrada para hacer predicciones o tomar decisiones basados en estos, en lugar de seguir estrictamente instrucciones estáticas y es empleado en un rango de tareas donde la programación explícita no es factible, como el filtrado de spam, reconocimiento óptico de caracteres, motores de búsqueda y visión por computadora.

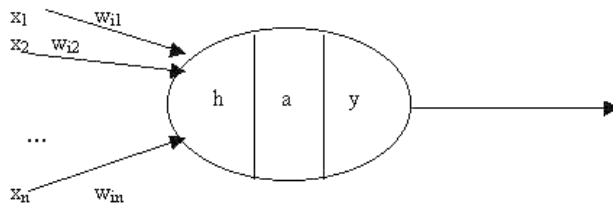
Las Redes Neuronales Artificiales son un paradigma del aprendizaje automático, inspirado en el funcionamiento del sistema nervioso animal. Es un sistema de interconexión de neuronas que colaboran entre sí para producir estímulos de salida.

Puede definirse como un grafo dirigido conformado por neuronas o elementos de proceso, pueden tener memoria local que junto a las entradas permiten emitir una señal de salida y conexiones unidireccionales indefinidas, que deben tener la misma señal.

3.1.1 Modelo de neurona artificial: El modelo de neurona de Rumelhart y McClelland desarrollado en 1986 es una unidad de cálculo que intenta modelar el comportamiento de una neurona animal real. Es la unidad básica y esencial para la construcción de redes neuronales artificiales.

El modelo define un elemento de proceso (EP) o neurona artificial, como un dispositivo que genera una única salida a partir de un conjunto o vector de entradas.

Figura 1. Modelo de Neurona Artificial.



Fuente: <http://avellano.usal.es/~lalonso/RNA/imagenes/modeloNA.jpg>

Este modelo consta de un conjunto o vector de entradas x , un conjunto de pesos sinápticos w_{ij} , una regla de propagación que suele ser una suma ponderada del producto escalar del vector de entrada y el vector de pesos, donde si las influencias excitadoras positivas dominan, se produce una señal positiva y transmite dicha señal a otras neuronas por su sinapsis de salida y puede ser expresada como,

$$h_i(t) = \sum w_{ij}x_j$$

También puede ser usada la distancia de Voronoi o de Mahalanobis, pero es más común el uso de la distancia euclídea entre ambos vectores:

$$h_i(t) = \sum (x_j w_i)^2$$

También cuenta con una función de activación $a_i(t)=f(a_i(t-1), h_i(t))$ que proporciona el estado de activación de la neurona en función del estado anterior y el valor post sináptico que no suele tener en cuenta el estado anterior de la neurona, sino sólo el potencial $h_i(t)$. Suele ser una función determinista y, casi siempre, continua y monótona creciente y con una función de salida $F_i(t)$ que proporciona la salida $y_i(t)$ en función del estado de activación. [6]

Figura 2. Funciones de Activación más comunes



Fuente: <http://avellano.usal.es/~lalonso/RNA/imagenes/funcionesActivacion.jpg>

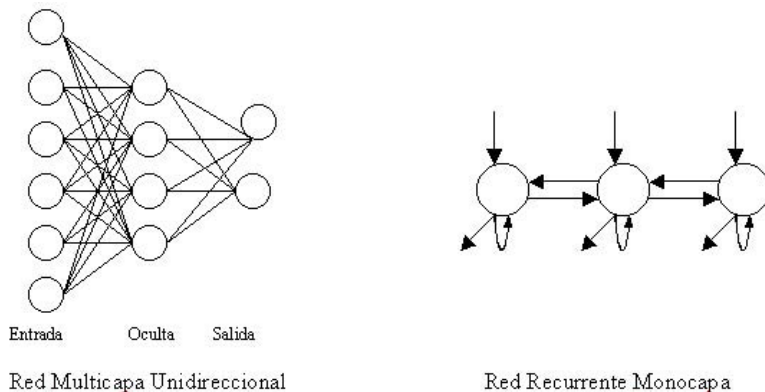
3.1.2 Arquitectura de las RNA: La arquitectura de una RNA es la estructura de conexiones de la red. Dado que las conexiones sinápticas son unidireccionales, la información se transmite en un solo sentido.

En general, las neuronas suelen agruparse en unidades estructurales llamadas capas. Dentro de una capa, las neuronas suelen ser del mismo. Estas redes están compuestas de 3 capas, las de entrada que son pasivas, es decir no modifican los datos sino que reciben información de entrada y la duplican en sus salidas, las capas ocultas y de salida son capas activas encargadas de evaluar los datos y proporcionar la respuesta de la red respectivamente.

Generalmente las conexiones se realizan entre neuronas de distintas capas, pero puede haber conexiones intracapa o *laterales* y conexiones de *realimentación* que siguen un sentido contrario al de entrada-salida.

En una estructura completamente conectada, cada valor de las capas de entrada es duplicado y enviado a las capas ocultas, donde son multiplicados por un peso predeterminado en cada nodo y posteriormente son sumados para producir un único número de salida, que es evaluado en una función sigmoide que limita su salida. Este proceso se repite para cada capa oculta hasta producir los valores de salida finales. [7]

Figura 3. Arquitectura de RNA



Fuente: <http://avellano.usal.es/~lalonso/RNA/imagenes/arquitecturaANS.jpg>

3.1.3 Aprendizaje de las RNA: El aprendizaje de la red es el proceso mediante el cual se refinan los pesos mediante un algoritmo de backpropagation, dados los resultados obtenidos en cada iteración comparado con los resultados deseados en el caso que sea aprendizaje supervisado y los datos estén etiquetados, buscando un conjunto de pesos que permitan obtener una salida lo más cercano posible a los valores reales de clasificación de los datos. En este caso la finalidad del algoritmo de aprendizaje es la minimización del error en los valores de salida por medio de una función de costo.

$$J(\theta) = \frac{1}{2} \sum_{t=1}^m (h_{\theta}(x^{(t)}) - (y^{(t)}))^2$$

Donde θ representa los pesos de la red, $h_{\theta}(x^{(t)})$ la hipótesis realizada y $y^{(t)}$ las etiquetas reales. Para hallar el valor óptimo de los pesos, es necesario que estos varíen con el tiempo, moviéndose en dirección de la pendiente decreciente de J más pronunciada. Estos valores pueden ser extraídos y almacenados para no tener que realizar el proceso de entrenamiento para hallarlos nuevamente.[8]

3.2 REDES NEURONALES CONVOLUCIONALES

Las redes neuronales convolucionales son redes igualmente conformadas por neuronas con un conjunto de pesos y sesgos aprendidos a lo largo del entrenamiento, con datos de entrada y salida en cada neurona. Sin embargo este tipo de redes asume que sus entradas son imágenes, buscando que sean más eficientes para dicho tipo de datos y utilizar una cantidad mínima de procesamiento a diferencia de las redes regulares que no funcionan tan bien con imágenes. Aunque su uso no se limita a imágenes, también son ampliamente usadas en reconocimiento de video y procesamiento de lenguaje.

Para el reconocimiento de imágenes éstas redes se componen de múltiples capas de pequeñas neuronas que procesan porciones de la imagen de entrada, llamadas campos receptores. Las salidas de éstas son apiladas para superponerse y obtener una mejor representación de la imagen original, proceso que se repite para cada capa. El apilamiento permite a éstas redes tolerar dicha traducción de las imágenes de entrada.

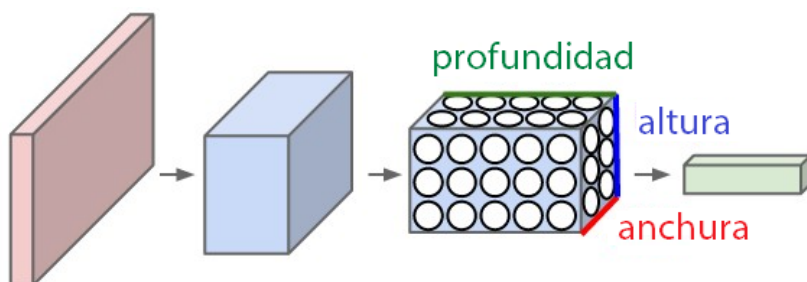
Usualmente suelen estar conformadas por varias capas de conexiones locales con sus respectivos pesos, capas de pooling y finalizar con capas completamente conectadas. El hecho de no estar completamente conectadas en sus capas ocultas, como sucede con las redes neuronales regulares, permite que estas redes sean más veloces, tengan menos parámetros y por tanto son más fáciles de entrenar. Para reducir el número de parámetros libres y mejorar la generalización se introduce una operación de convolución en pequeñas regiones de la entrada. Una gran ventaja de las redes convolucionales es el uso de pesos compartidos en las capas convolucionales, es decir que el mismo filtro o banco de pesos es usado para cada pixel de la capa, reduciendo la huella de memoria y mejora el rendimiento. Comparado con otros algoritmos de clasificación de imágenes, las redes neuronales convolucionales usan relativamente poco pre procesamiento. Esto quiere decir que la red es responsable de aprender los filtros que deben ser programados manualmente en los algoritmos tradicionales. La falta de dependencia de conocimiento a priori y de esfuerzo humano en el diseño de características es una gran ventaja de las CNN. [3]

3.2.1 Características distintivas: El tradicional modelo de los perceptrones multicapas fue exitosamente usado para reconocimiento de imágenes. No se adaptan bien a las imágenes, dada la conectividad de todos los nodos entre capas y no toma en cuenta la estructura espacial de los datos, tratando los píxeles de entrada cercanos en el mismo nivel que los muy lejanos. Una conectividad completa entre neuronas es un desperdicio en el framework de reconocimiento de imágenes y el gran número de parámetros lleva rápidamente al sobre ajuste de éstos.

Las CNN están diseñadas para emular el comportamiento de la corteza visual, mitigan los retos de la arquitectura anterior al explotar la fuerte correlación espacial local presente en imágenes naturales. Opuesto al perceptrón multicapa, las redes convolucionales tienen las siguientes características distintivas:

- Volúmenes 3D de neuronas: Las capas de una CNN tienen neuronas en arreglos tridimensionales, a lo alto, ancho y profundo. Las neuronas dentro de la capa solo están conectadas a una pequeña región de la capa anterior, llamada campo receptivo. Distintas clases de capas son agrupados para formar la arquitectura de la red convolucional.

Figura 4. Capas de una CNN es un arreglo tridimensional



Fuente: <http://cs231n.github.io/assets/cnn/cnn.jpeg>

- Conectividad local: las CNN explotan la correlación espacial al reforzar un patrón de conectividad local entre neuronas de capas adyacentes. La arquitectura asegura que los filtros aprendidos produzcan la mejor respuesta a un patrón de entrada local. Apilar tantas capas lleva a que los filtros no lineales se conviertan incrementalmente en filtros globales, permitiendo que la red cree buenas representaciones de pequeñas partes

de la entrada y posteriormente ensamble representaciones de áreas más grandes de éstas.

- **Pesos Compartidos:** Cada filtro es replicado a través de todo el campo visual. Dichas unidades replicadas comparten una misma parametrización y forman un mapa de características. Esto quiere decir que todas las neuronas en una capa convolucional, detecta exactamente la misma característica. Las unidades replicadas entonces permiten que las características sean detectadas a pesar de su posición en el campo visual, constituyendo así la propiedad de la traducción invariante. [6]

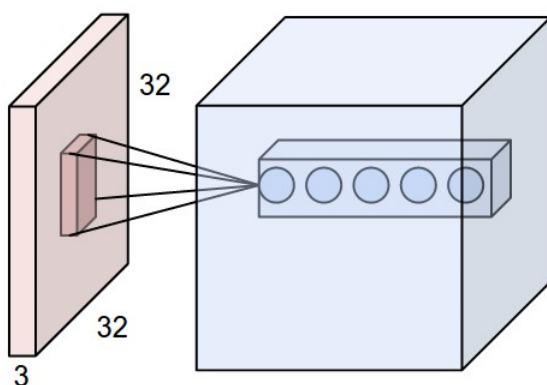
La disolución del gradiente es un problema bastante común en aprendizaje de máquina y en especial en las redes neuronales al tener métodos de aprendizaje basados en gradientes. En éstos, cada uno de los pesos de la red recibe una actualización proporcional al gradiente de la función de error con respecto al peso de cada una de las iteraciones del entrenamiento. El gradiente tiende a desaparecer por su disminución exponencial en las capas iniciales de la red. Para poder solucionar este problema, es posible pre entrenar la red, mejorando su acierto en la generalización. Dado que la red es expuesta a una gran cantidad de datos, los parámetros de pesos son tratados en un espacio diferente donde es más factible representar la distribución de los datos en general, en lugar que se sobre ajusten a un subconjunto de la distribución de los datos. Al realizar un pre entrenamiento, es posible actualizar los pesos hasta cierto punto, disminuyendo el efecto de la disolución del gradiente, necesitando menos iteraciones para alcanzar el valor deseado.

3.2.2 Arquitectura Una CNN consiste de un número de capas convolucionales y submuestreo, seguidas opcionalmente por capas completamente conectadas. La entrada a una capa convolucional es una imagen de dimensiones $m \times m \times r$. La capa convolucional tiene k filtros o kernel de tamaño $n \times n \times q$, donde n es más pequeño que la dimensión de la imagen, q puede ser igual o menor al número de canales r de ésta y puede variar para cada kernel. El tamaño de los filtros da lugar a la estructura localmente conectada convolucionada con la imagen para producir k mapas de características de tamaño $m-n+1$. Cada mapa es posteriormente submuestreado con agrupamiento medio o máximo en $p \times p$ regiones contiguas donde p tiene un rango de 2 para imágenes pequeñas y no suele ser mayor que 5 para entradas más grandes. Suelen formarse con capas convolucionales, de agrupamiento, ReLu, completamente conectada y de pérdida. [8]

Capa convolucional: Es el bloque de construcción central de una CNN. Los parámetros de la capa consisten en un conjunto de filtros que pueden ser aprendidos, llamados kernel, que tienen un pequeño campo receptivo pero se extienden a través de la totalidad de la profundidad del volumen de entrada. Durante el pase hacia adelante, cada filtro es convolucionado a lo alto y ancho del volumen de entrada, computando el producto punto entre las entradas del filtro y la entrada produciendo un mapa de activación de 2 dimensiones en dicho filtro. Gracias a esto, los filtros son aprendidos por la red y pueden ser activados al detectar una característica en los datos de entrada.

La salida de una capa de convolución está formada por todos los mapas de activación aprendidos. Cada entrada en el volumen de salida también puede interpretarse como una salida de una neurona que mira una pequeña región en la entrada y comparte parámetros con otra en el mismo mapa de activación. [4]

Figura 5. Neuronas de una capa convolucional conectada a su campo receptivo.



Fuente: <http://cs231n.github.io/assets/cnn/depthcol.jpeg>

Las redes convolucionales explotan la correlación espacial local al reforzar un patrón de conectividad local entre neuronas de capas adyacentes, cada neurona está conectada solo a una pequeña región del volumen de entrada. El alcance de ésta es un hiper parámetro llamado campo receptivo de la neurona. Las conexiones son locales en el espacio, siempre se extienden a lo largo y ancho de la profundidad total del volumen de entrada. Dicha arquitectura asegura que los filtros aprendidos produzcan la respuesta más fuerte a un patrón de entrada espacialmente local. Esto es conocido como la conectividad local.

El tamaño del volumen de salida de una red convolucional es controlado por 3 hiper parámetros, profundidad, paso y rellenado con ceros.

- La profundidad del volumen de salida controla el número de neuronas en la capa que conecta a la misma región del volumen de entrada. Todas éstas aprenden a activarse para diferentes características en la entrada.

- El paso controla cómo se distribuyen las columnas de profundidad alrededor de las dimensiones espaciales. Cuando el paso es 1, una nueva columna de neuronas se asigna a una posición espacial a una unidad espacial de distancia. Esto lleva a la rápida superposición de campos receptivos entre columnas y a una gran cantidad de volúmenes de salida. Inversamente, al usar mayores pasos los campos receptivos se solapan menos y el volumen de salida tendrá menos dimensionalidad espacial.

- En ocasiones es conveniente rellenar la entrada con ceros en el borde del volumen. El tamaño de este rellenado es el tercer hiper parámetro. Permite controlar el tamaño de los volúmenes de salida, aunque en ocasiones es deseable conservar el tamaño espacial exacto.

El tamaño espacial del volumen de salida puede ser computado como una función del tamaño del volumen de entrada W , el tamaño del campo receptor de las neuronas de la capa convolucional F , el paso aplicado S y la cantidad de rellenado con ceros en el borde P . La fórmula para calcular cuántas neuronas cabrán en el volumen dado es $\frac{W-F+2P}{S+1}$. El resultado de esta operación siempre será un entero, de no ser así, indica asignaciones erróneas y por tanto las neuronas no pueden apilarse apropiadamente y de forma simétrica en el volumen de entrada. En general asignar el rellenado en $P = \frac{F-1}{2}$ cuando el paso es 1, asegura que el volumen de entrada y de salida tenga el mismo tamaño espacial.

El intercambio de parámetros es usado para controlar el número de parámetros libres. Asume que con una porción bidimensional a profundidad, las neuronas en cada porción de profundidad usarán los mismos pesos y sesgo. Como todas las neuronas a una misma profundidad comparten parametrización, el paso de la información hacia la siguiente capa en cada porción a profundidad de la capa, puede ser computada como una convolución de los pesos de las neuronas con el volumen de entrada. Los pesos convolucionados con la entrada son conocidos como filtros. El resultado de dicha operación es un mapa de activación y los conjuntos de éstos para cada filtro diferente son apilados juntos en la dimensión de profundidad para producir el volumen de salida. Compartir parámetros contribuye a la invariancia en la traducción de la arquitectura de las redes neuronales convolucionales.

Capa de Agrupamiento: Es una capa altamente importante, pues se encarga de disminuir el tamaño después de extraer las características por medio de la convolución. El agrupamiento particiona la imagen de entrada en un conjunto de rectángulos sin superposición y para cada uno de estos, devuelve el valor máximo. La posición exacta de la característica encontrada no es tan importante como la relativa a otras características. La función de la capa de agrupamiento es reducir la cantidad de parámetros y de computación en la red al disminuir progresivamente el tamaño espacial de la representación consiguiendo controlar también el sobreajuste a los datos. Es común insertar capas de agrupamiento entre capas de convolución. Ésta provee una invariancia en la traducción de los datos, lo que significa que una misma característica estará activa aún si la imagen es sometida a traducciones y reducciones. Se desea que las características obtenidas sean invariantes.

El agrupamiento máximo es actualmente el más usado en el desarrollo de las redes, pues ha demostrado los mejores resultados en la práctica. Su salida es el valor máximo de cada región. También existen otras funciones de agrupamiento como la media y la norma L2. Dada la agresiva reducción en el tamaño de la representación, se tiende a usar filtros pequeños.

Capa ReLu: Es una capa de neuronas que aplica la función de activación no saturante $f(x) = \max(0, x)$ Aumenta las propiedades no lineales de la función decisoria y de la red en general sin afectar los campos receptivos de la capa convolucional. Entre otras funciones usadas para aumentar la no linealidad se encuentran la tangente hiperbólica $f(x) = \tanh(x)$ y la función sigmoide $f(x) = 1/(1+e^{-x})$ Comparada a otras funciones, es preferible el uso de ReLu dando un entrenamiento más rápido de la red sin causar cambios significativos en la precisión de la generalización. [5]

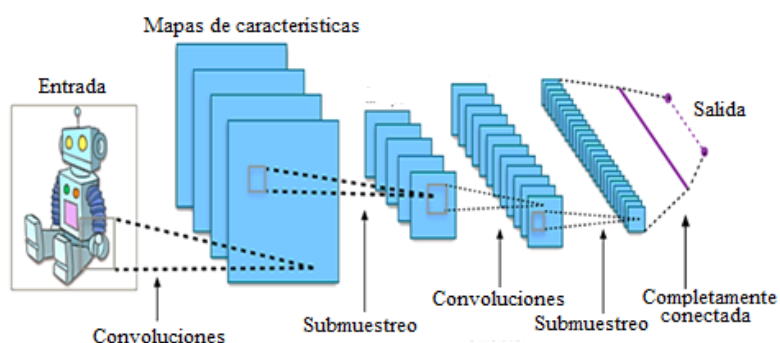
Capa completamente conectada: Después de varias capas convolucionales y de agrupamiento máximo, se realiza el razonamiento de alto nivel por medio de capas completamente conectadas. Las neuronas de ésta, tienen conexiones completas a todas las activaciones en la capa inmediatamente anterior, por tanto sus activaciones pueden ser computadas con una multiplicación de matrices seguida por desplazamiento del sesgo.

Capa de pérdida: Especifica la penalización hecha por la red en la diferencia entre el valor predicho y valor real de la etiqueta de los datos. Las funciones softmax

son utilizadas en la predicción de clases únicas y mutuamente excluyentes entre sí, mientras que la entropía sigmoide cruzada se usa en predicciones en valores continuos entre 0 y 1 y la Euclídeana para la regresión de etiquetas con valores reales.

Una arquitectura típica de redes neuronales convolucionales está conformada por algunas capas ReLu, seguidas por capas de agrupamiento, repitiendo este patrón, hasta que el valor de entrada haya sido unido espacialmente hasta alcanzar un tamaño mucho menor, que pueda considerarse pequeño. Llegado un punto, es natural, hacer la transición a una capa completamente conectada, que mostrará el valor o valores de salida correspondientes. [1]

Figura 6. Arquitectura típica de una CNN



Fuente: https://upload.wikimedia.org/wikipedia/commons/6/63/Typical_cnn.png

3.2.3 Métodos de regularización

- Dropout: Es uno de los métodos más populares pues ha demostrado reducir el sobreajuste de los datos y consiste en poner la salida de cada neurona oculta en la capa elegida a 0 para evitar entrenar con todos los nodos y datos de entrada. Las capas completamente conectadas tienden a estar sobre ajustadas, pues conservan la mayoría de los parámetros. Este método permite que las características aprendidas sean más robustas para una mejor generalización. En cada etapa del entrenamiento, nodos individuales son eliminados de la nueva red con una probabilidad de $1-p$ o se mantienen con probabilidad p , produciendo una red reducida, eliminando

también todos los nodos entrantes o salientes de un nodo eliminado y se entrena con la red obtenida. Posteriormente se reinsertan los nodos con sus pesos originales.

En las etapas de entrenamiento, la probabilidad que un nodo en las capas medias sea mantenido suele ser de 0.5.

- DropConnect: Fue usado en conjuntos de datos pequeños y su desempeño es bueno, aunque no tanto como el de Dropout y a diferencia de este, las neuronas con pesos puestos en cero son elegidos aleatoriamente y cada unidad recibe la entrada de un subconjunto aleatorio de la capa inmediatamente anterior.
- Stochastic pooling: Se realizan procedimientos estocásticos para reemplazar las operaciones determinísticas de agrupamiento (o pooling) para buscar un mejor ajuste a las redes convolucionales. La respuesta del mapa de agrupamiento es seleccionada por el muestreo de una distribución multinomial formada de las activaciones de cada región de agrupamiento. Más precisamente, se inicia computando las probabilidades p para cada región j al normalizar las actividades de cada una:

$$p_i = \frac{a_i}{\sum_{k \in R_j} a_k}$$

Aleatoriamente se eligen activaciones en cada región de pooling de acuerdo a una distribución multinomial, dadas las actividades en dicha región. Entonces la activación agrupada es a_l :

$$s_j = a_l \text{ donde } l \sim P(p_1, \dots, p_{|R_j|})$$

Las muestras para cada región agrupada en cada capa y para cada ejemplo de entrenamiento son independientes entre sí. [8]

- Datos Artificiales: El grado de sobre ajuste en la curva de aprendizaje de un modelo está determinado por su poder y la cantidad de datos de entrenamiento que recibe, proveer más datos de entrenamiento a una red convolucional, puede ayudar a reducirlo. La idea es proporcionar más imágenes a la red, pero usualmente en el proceso se utilizan todas las disponibles, entonces en aras de disminuir dicho sobreajuste se perturban los datos originales, para obtener nuevos ejemplos.

- **Tamaño de la red:** Limitar el número de unidades ocultas y parámetros libres en la red, es la forma más sencilla de prevenir su sobreajuste, restringiendo directamente el poder predictivo de la red y reduciendo la complejidad de la función que puede realizar en los datos.
- **Decadencia del peso:** consiste en adicionar un error proporcional a la suma de los pesos o la magnitud al cuadrado del vector de los pesos. El nivel de complejidad puede ser reducido al incrementar la constante de proporcionalidad, aumentando la penalidad para vectores con pesos altos.
 - **Regularización L2:** Se implementa penalizando el cuadrado de la magnitud de todos los parámetros directamente en el objetivo. Penaliza altamente los vectores de peso que difieren mucho de la media de éstos y debido a las interacciones multiplicativas entre pesos y entradas, tiene la propiedad de incentivar el poco uso de todas las entradas, en lugar de unas pocas en grandes cantidades.
 - **Regularización L1:** Tiene la propiedad de hacer los vectores de los pesos sean vectores dispersos durante la optimización, es decir usa un subconjunto disperso de sus entradas más importantes y se vuelve casi invariante a las entradas ruidosas. Es posible combinar las regularizaciones L1 y L2, bajo el nombre de regularización elástica.
- **Restringir la norma máxima:** Es posible aplicar un límite superior absoluto a la magnitud del vector peso para cada neurona y usar gradientes descendentes proyectados para reforzar la restricción. En la práctica, corresponde a la actualización de parámetros y luego realizar la restricción fijando el vector de pesos w de cada neurona para satisfacer que [2]

$$\|\bar{w}\|_2 < c$$

3.3 PLANCTON

Son organismos acuáticos principalmente microscópicos incapaces de nadar, que flotan en aguas saladas o dulces y son transportados por las corrientes. El término plancton, no define al organismo, sino su forma de vida. Existen dos grandes grupos de plancton, clasificados según su estilo de vida, si es similar a las plantas se denomina fitoplancton y si es similar a los animales, zooplancton. El plancton vegetal está siempre cerca de la superficie del agua, pues necesita luz para realizar la fotosíntesis. En cambio el zooplancton está siempre en movimiento, de arriba hacia abajo, completando un ciclo diario con un recorrido de entre 100 a 400 metros. Están cerca de la superficie de noche para alimentarse, y más abajo durante el día para escapar de las fuertes radiaciones solares.

La mayoría de las especies son transparentes con una cierta iridación, y presentan colores sólo al microscopio. Las especies superficiales son azuladas, y las otras rojizas. Algunas emiten luminiscencia, como la noctiluca.

La mayoría de las especies del plancton mide menos de un milímetro, otras en cambio son más grandes, como los sifonóforos, ctenóforos y medusas acalefas.

3.3.1 Importancia del plancton: El plancton puede ser productor, consumidor o ambos. Juega un papel importante en la cadena alimentaria acuática y son fuente de alimento para organismos más grandes. El fitoplancton son productores que fotosintetizan, es decir, convierten la energía solar, dióxido de carbono y nutrientes presentes en el agua, en comida para sí y contribuyen con hasta un 50% del oxígeno atmosférico. El zooplancton es una conexión muy importante entre productores primarios y niveles tróficos más altos.

3.3.2 Proliferación de plancton: Generalmente las poblaciones de fitoplancton se verifican constantemente porque su crecimiento está limitado por factores como luz o nutrientes presentes en el agua. Sin embargo, cuando uno de estos factores limitantes es abundante, una proliferación de plancton, caracterizada por un gran aumento en su población, puede ocurrir. Éstas suceden naturalmente en primavera, cuando aumenta la luz solar, la temperatura y los nutrientes de tierra firme son traídos al agua y suelen ser por la rápida reproducción de una sola especie de fitoplancton y dependiendo qué tan grande sea este fenómeno, puede llegar a cambiar el color del agua.

Figura 7. Imagen satelital de una proliferación de plancton, color verde claro en la costa de Islandia. Cortesía del Observatorio Terrestre de la NASA.



Fuente: http://d32ogoqmya1dw8.cloudfront.net/images/eslabs/carbon/phytoplankton_bloom_norway_744.jpg

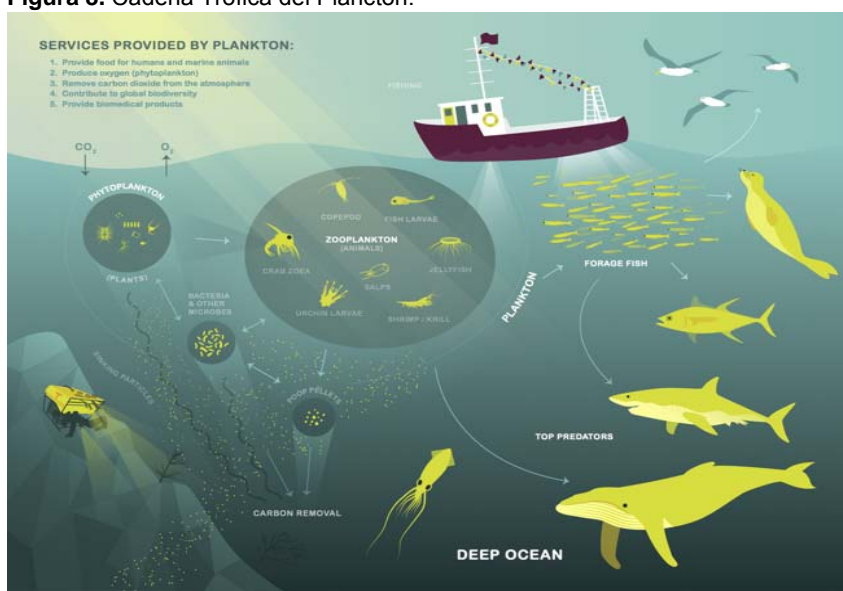
Las crecientes proliferaciones de plancton en las últimas décadas preocupan grandemente a los investigadores, pues impactan los ecosistemas acuáticos y pueden ser perjudiciales, en especial cuando las especies en aumento son tóxicas o se reproducen en magnitudes extremas. En este último caso, el exceso de fitoplancton se hunde al fondo del agua, donde son descompuestos por bacterias, que consumen oxígeno al descomponerlas, lo que puede conllevar a bajos niveles de oxígeno disuelto, resultando en una zona muerta donde el agua no puede soportar la vida. Las zonas muertas pueden tener grandes impactos en los ecosistemas acuáticos, por lo que aún se investigan los efectos que puede tener a nivel de organismos individuales y grupos de éstos.

3.3.3 Clasificación: Aunque tradicionalmente se ha subdividido el plancton en fitoplancton y zooplancton, según las clasificaciones más recientes esta distinción no parece apropiada, ya que entre los organismos autótrofos se incluyen los vegetales, algunos protistas y bacterias, y entre los heterótrofos están los animales, otros protistas y bacterias. No obstante, esta clasificación sigue utilizándose comúnmente.

El fitoplancton es el alimento del zooplancton. Éste, sirve al mismo tiempo como alimento a equinodermos, crustáceos y peces en estado larvario. Estas larvas al crecer sirven como alimento a bancos de pequeños peces que a su vez alimentan a grandes planctívoros, como las ballenas o los tiburones ballena, y peces

mayores que alimentan, en varios eslabones sucesivos, a los grandes depredadores oceánicos, como son los cetáceos carnívoros, los tiburones, los atunes o los peces espada. En proporción, una tonelada de estos últimos habrá requerido, para su existencia y desarrollo, cinco mil toneladas de fitoplancton, como parte de lo que se denomina cadena trófica.

Figura 8. Cadena Trófica del Plancton.



Fuente: <https://kaggle2.blob.core.windows.net/competitions/kaggle/3978/media/Plankton-Diagram3-lg.png>

- **Zooplankton:** Constituido por todos los consumidores que constituyen en su gran mayoría a productores secundarios y terciarios. Este grupo está constituido por organismos generalmente microscópicos adultos y sus fases larvares conocido como holoplancton, y por las fases larvares de otros organismos que en forma adulta habitan los fondos acuáticos o la columna de agua, contrarrestando el movimiento de las corrientes. Algunos de los grupos de organismos más abundantes y característicos del zooplankton son los copépodos, cladóceros, rotíferos, cnidarios, quetognatos, eufáusidos y las larvas de los peces que por su relevancia socioeconómica de los organismos juveniles y

adultos generalmente estudian y describen con el término "ictioplancton". Al igual que el fitoplancton, dependiendo del ambiente en que se encuentren, ya sea dulceacuícola o marino, cada uno de los grupos o especies del zooplancton variará su diversidad y abundancia. Un componente del zooplancton relativamente menos estudiado son sus parásitos que constituyen una diversidad varios órdenes de magnitud mayor que los mismos organismos fitoplanctónicos y zooplanctónicos ya que cada organismos que existe en el planeta es propenso a infestarse o infectarse por múltiples parásitos.

- *Fitoplancton*: El plancton vegetal, se desarrolla en las aguas costeras del mar con luz solar y sales minerales abundantes de hasta 30 m de profundidad, dado que elaboran su alimento por fotosíntesis. Constituyen el alimento del zooplancton y producen el 50 % del oxígeno molecular necesario para la vida terrestre. Los organismos que más abundan en el fitoplancton son las cianobacterias y las diatomeas, unas algas doradas unicelulares. También encontramos a los dinoflagelados, responsables de las mareas rojas. Base de la cadena trófica marina, el fitoplancton ha experimentado un significativo descenso debido al aumento de la radiación ultravioleta. Se ha observado que bajo el agujero de la capa de ozono en la Antártida la productividad del fitoplancton decreció entre el 6 % y el 12 %.

3.3.4 Otras clasificaciones

- Según su tamaño

Diversos autores han realizado una clasificación del plancton por su tamaño, aunque puede considerarse "artificial" al basarse inicialmente en la luz de malla con la que se hacían las capturas, y no se ha llegado a un acuerdo definitivo. Una de las clasificaciones más utilizadas es la siguiente:

- Femtoplancton (0,02 - 0,2 μm): virus, bacterias
- Picoplancton (0,2 - 2 micras): bacterias, cianobacterias, prasinofices
- Nanoplancton (2 – 20 micras): fitoflagelados
- Microplancton (20 – 200 micras): diatomeas, peridíneas, tintinados, radiolarios, ostrácodos, cladóceros
- Macroplancton (0,2 – 2 mm): medusas, alevines
- Megaloplancton (mayor de 2 mm): medusas, colonias de tunicados.

- Según su distancia
 - Según su distancia con respecto a la costa
 - Plancton nerítico: es el que se encuentra sobre la plataforma continental.
 - Plancton oceánico: es el que se encuentra fuera de la plataforma continental.
 - Según su profundidad
 - Epiplancton: 0 - 50 metros.
 - Subplancton: 50 - 200 metros.
 - Infraplancton: 200 - 500 metros.
 - Batiplancton más de 500 metros bajo la superficie

Los oceanógrafos del Centro de Ciencia Marina Hatfield de la Universidad Estatal de Oregon, han clasificado el plancton en fitoplancton y zooplancton, y según su taxonomía en 121 clases diferentes, que incluyen diversos tipos de copépodos, crustáceos, ctenóforos, detritos, larvas e hidromedusas.

Los oceanógrafos de la universidad de Hatfield, han clasificado el plancton según sus características antropológicas en 121 clases diferentes, incluyendo 3 clases de plancton sin clasificación denominadas manchas desconocidas, varas desconocidas y desconocidos, sin clasificación.

Tabla 1. Clasificación taxonómica del Plancton

acantharia_protist	echinoderm_larva_pluteus_typeC	protist_dark_center
acantharia_protist_big_center	echinoderm_larva_pluteus_urchin	protist_fuzzy_olive
acantharia_protist_halo	echinoderm_larva_seastar_bipinnaria	protist_noctiluca
amphipods	echinoderm_larva_seastar_brachiolaria	protist_other
appendicularian_fritillaride	echinopluteus	protist_star
appendicularian_s_shape	ephyra	pteropod_butterfly
appendicularian_slight_curve	euphausiids	pteropod_theco_dev_seq

appendicularian_straight	Euphausiids_young	pteropod_triangle
artifacts	fecal_pellet	radiolarian_chain
artifacts_edge	fish_larvae_deep_body	radiolarian_colony
chaetognath_non_sagitta	fish_larvae_leptocephali	shrimp_caridean
chaetognath_other	fish_larvae_medium_body	shrimp-like_other
chaetognath_sagitta	fish_larvae_myctophids	shrimp_sergestidae
chordate_type1	fish_larvae_thin_body	shrimp_zoea
copepod_calanoid	fish_larvae_very_thin_body	siphonophore_calycophoran_abyllidae
copepod_calanoid_eggs	heteropod	siphonophore_calycophoran_rocketship_adult
copepod_calanoid_eucalanus	hydromedusae_aglaura	siphonophore_calycophoran_rocketship_young
copepod_calanoid_flatheads	hydromedusae_bell_and_tentacles	siphonophore_calycophoran_sphaeronectes
copepod_calanoid_frillyAntennae	hydromedusae_h15	siphonophore_calycophoran_sphaeronectes_stem
copepod_calanoid_large	hydromedusae_haliscera	siphonophore_calycophoran_sphaeronectes_young
copepod_calanoid_large_side_antennatucked	hydromedusae_haliscera_small_sideview	siphonophore_other_parts
copepod_calanoid_octomoms	hydromedusae_liriope	siphonophore_partial
copepod_calanoid_small_longantennae	hydromedusae_narcondark	siphonophore_physonect
copepod_cyclopoid_copilia	hydromedusae_narcomedusae	siphonophore_physonect_young
copepod_cyclopoid_oithona	hydromedusae_narco_young	Stomatopod
copepod_cyclopoid_oithona_eggs	hydromedusae_other	tornaria_acorn_worm_larvae
copepod_other	hydromedusae_partial_dark	trichodesmium_bowtie
crustacean_other	hydromedusae_shapeA	trichodesmium_multiple
ctenophore_cestid	hydromedusae_shapeA_sideview_small	trichodesmium_puff
ctenophore_cydippid_no_tentacles	hydromedusae_shapeB	trichodesmium_tuft
ctenophore_cydippid_tentacles	hydromedusae_sideview_big	trochophore_larvae

ctenophore_lobate	hydromedusae_solmaris	tunicate_doliolid
decapods	hydromedusae_solmund ella	tunicate_doliolid_nurse
detritus_blob	hydromedusae_typeD	tunicate_partial
detritus_filamentus	hydromedusae_typeD_be ll_and_tentacles	tunicate_salp
detritus_other	hydromedusae_typeE	tunicate_salp_chains
diatom_chain_string	hydromedusae_typeF	unkown_blobs_and_smudges
diatom_chain_tube	invertebrate_larvae_other _A	unkown_sticks
echinoderm_larva_pluteus_ brittlestar	invertebrate_larvae_other _B	unkown_unclassified
echinoderm_larva_pluteus_ early	jellies_tentacles	

3.3.5 Distribución del plancton: Los movimientos del plancton en el agua han de entenderse como desplazamientos verticales y no horizontales sobre una columna de agua. La mayor producción del plancton vegetal se sitúa en los primeros 50 metros de la columna vertical del agua. Por debajo de esta profundidad, encontramos fitoplancton hasta unos 200 metros, aunque en menor cantidad y calidad y con la particularidad de que, con la ayuda de las corrientes, debe subir más arriba a renovarse.

A diferencia del fitoplancton, que se sitúa sobre una capa poco profunda del agua para captar la luz solar, el zooplancton mantiene un movimiento vertical mucho más amplio. Se puede encontrar zooplancton hasta en las grandes fosas marinas, sin embargo su rango más óptimo de acción, al igual que el fitoplancton son los 50 primeros metros. Entre los 50 y los 600 metros de profundidad aparece una zona con abundante plancton animal y, a partir de aquí va disminuyendo en mayor proporción.

Muchos miembros del plancton animal, por la noche se desplazan mediante flagelos hacia las capas superiores para alimentarse del plancton vegetal. Por el día la mayor proporción de organismos se suele quedar entre los 200 y los 600 metros para no ser dañados por los rayos del sol o desciende a capas más profundas para alimentarse. Hay especies que permanecen siempre en esta capa.

Se ha considerado que el zooplancton que vive en capas muy profundas se alimentaba solamente de los restos caídos de los animales que vivían en capas superiores. Estudios actuales parecen constatar que la riqueza alimentaria de los

detritos depositados en el fondo es tan pequeña que el plancton animal se ve obligado a subir a niveles no tan profundos para alimentarse de zooplancton que, a su vez, también sube más arriba para alimentarse de fitoplancton. Al mismo tiempo, el zooplancton que vuelve a las grandes profundidades sirve de alimento a otro zooplancton que no se mueve de su lugar.

Debe entenderse la cadena trófica acuática como una columna vertical cuyo mayor nivel de eficacia aparece en la capa superior o superficie y que va disminuyendo progresivamente a medida que bajamos a capas más profundas. [12]

4. ESTADO DEL ARTE

Tradicionalmente, la clasificación del plancton se lleva a cabo mediante procesos manuales y son contados gracias a las muestras tomadas usando un microscopio, por lo que el proceso toma meses, incluso años. Esto no permite que los estudios se realicen a gran escala o con determinada periodicidad.

En un intento de superar las limitaciones, en 1992 fue desarrollada una Video Grabadora de Plancton, que por primera vez capturaba imágenes individuales de éste, permitiendo el procesamiento de grandes volúmenes de datos, abriendo puertas a software y hardware para su reconocimiento y diversos tipos de algoritmos de reconocimiento de patrones. [2]

El primer proyecto para la automatización de dicho proceso fue realizado en 1998, por Tang Xiaou y su equipos, quienes aprovechan los momentos invariantes de las imágenes, los descriptores de Fourier, la extracción de características y la reducción de estas por medio de la transformación Karhunen-Loeve para su procesamiento en el clasificador de aprendizaje de cuantización vectorial, que consideraron suficientemente robusto dada la distribución no lineal de los datos, como lo es el caso de las imágenes de plancton obtenidas. [9]

En 2007 fue desarrollado un robot sumergible llamado FlowCybot, que captura imágenes y clasifica de 22 clases de fitoplancton con un 88% de eficacia, combinando los métodos de máquinas de soporte vectorial y clasificación de características. [8]

El instituto alemán Fraunhofer, desarrolló un proyecto para la segmentación y clasificación semiautomática del zooplancton. Dicho software se encarga del procesamiento de imágenes y los métodos de clasificación. [11]

Figura 9. Vista externa de FlowCybot



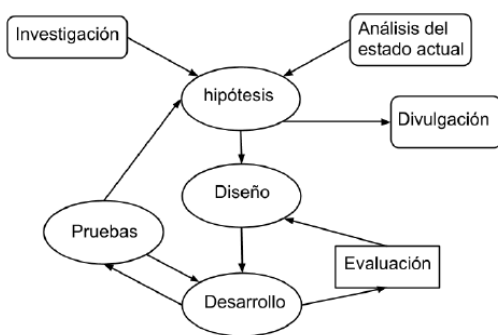
5. METODOLOGÍA

El presente proyecto fue desarrollado por medio una metodología experimental ágil, para lograr el análisis, diseño e implementación del software para la clasificación automática de plancton según su taxonomía.

El proyecto surge como resolución a un problema existente en la comunidad oceanográfica y un reto propuesto en la plataforma Kaggle patrocinado por ellos y la empresa Booz Allen Hamilton, titulado National Data Science Bowl. El Centro de Ciencia Marítima de la universidad de Óregon proporcionó los datos utilizados, que constan de 33,336 imágenes para entrenamiento de resolución variable, divididas en 121 clases diferentes, que se fijarán a una resolución de 256 x 256 al momento de subirlas al framework Digits, proceso que realiza automáticamente para garantizar su correcto funcionamiento y 130,400 imágenes sin clasificación para pruebas.

Se inició con el procesamiento de las imágenes, descrito en la sección de desarrollo, seguido por la creación de un diseño de red convolucional para llevar a cabo la tarea. Basada en los resultados obtenidos de esta primera prueba, se crearon e implementaron más diseños de red, variando la cantidad de capas y parámetros que se logró un buen desempeño. Sin embargo, como sólo se tenía un grupo de imágenes con un preprocesamiento específico, se procedió a procesar las imágenes de diversas formas para probar su desempeño en las redes con buenos resultados de clasificación.

Figura 10. Metodología de trabajo, metodología ágil experimental.



5.1 ETAPA DE BOSQUEJO DEL DISEÑO DE LA SOLUCIÓN

Se iniciará el proceso de investigación en redes neuronales convolucionales para la clasificación de imágenes. Los diseños de las redes existentes serán ejecutados con los datos a clasificar, para examinar los diseños que más favorezca los datos, entender su arquitectura y hacer un bosquejo en papel de lo que parezca mejorar el desempeño.

5.2 ETAPA DE DESARROLLO DEL PROTOTIPO

Se procesarán las imágenes en busca de una mejor representación y se desarrollará un primer prototipo de red neuronal convolucional para la clasificación del plancton, según el desempeño visto en las redes existentes.

5.3 ETAPA DE IMPLEMENTACIÓN Y PRUEBAS

Con los resultados del primer prototipo se seguirán creando redes convolucionales con variación en sus diseños, dependiendo del porcentaje de acierto general obtenido en cada una.

5.4 ETAPA DE DOCUMENTACIÓN

Una vez obtenidos resultados contundentes de la relación entre la arquitectura de la red y el desempeño de la misma, se procederá a documentar los diseños realizados, su porcentaje de acierto y su evolución.

6. DESARROLLO DEL PROYECTO

En este capítulo se presenta una descripción del diseño e implementación de las redes neuronales empleadas para la clasificación y el procesamiento de las imágenes de plancton por medio del uso de una metodología de desarrollo ágil.

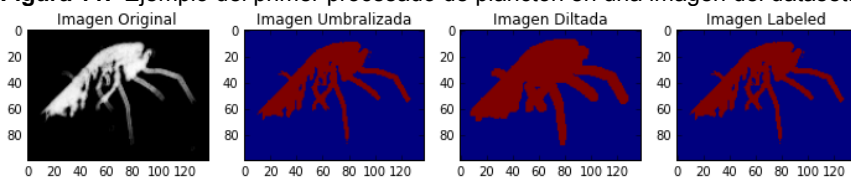
6.1 FASE 1.

6.1.1 Procesamiento de imágenes Fueron provistas 30,336 imágenes pertenecientes a 121 clases diferentes en un dataset no balanceado, es decir cuyas clases no tienen una cantidad similar de datos. En promedio las imágenes tienen una resolución de 73 x 66.

Inicialmente se realizó una revisión de los datos y la forma como estaban agrupados al igual que las librerías de procesamiento de imágenes de python, en este caso se utilizó la librería skimage.

Las imágenes fueron umbralizadas para convertirlas en binarias y extraer los píxeles que representan el objeto en sí, aunque esto puede generar pequeñas formas indeseables que no son realmente parte del objeto en sí, por tanto, es necesario realizar una operación morfológica. Se eligió el dilatado y posteriormente por medio de un filtro label, se unieron los píxeles vecinos con iguales valores de intensidad, buscando recuperar la información perdida por el microscopio.

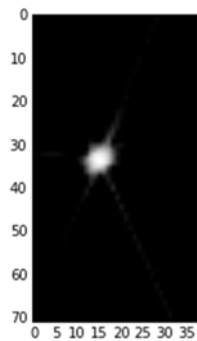
Figura 11. Ejemplo del primer procesamiento de plancton en una imagen del dataset.



Como siguiente método de procesamiento se les aplicó un filtro de la transformada rápida de Fourier, obteniendo los datos en un dominio diferente para evaluar el

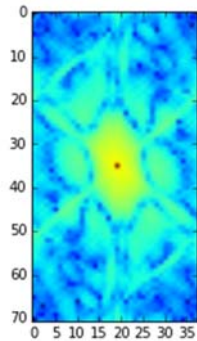
nivel de abstracción en la extracción de características, eliminando las frecuencias insignificantes manteniendo los detalles, que para el caso de nuestros datos es importante pues se suelen filtrar manchas y errores que no pertenecen a la imagen.

Figura 12. Ejemplo del segundo filtro aplicado, la transformada de Fourier.



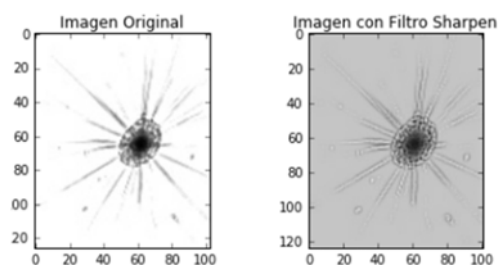
Posteriormente se usó un logaritmo en base 10 a las imágenes previamente umbralizadas y dilatadas, para la compresión dinámica del rango de las mismas. Éste suprime los cambios pequeños de la imagen y resalta donde encuentra un cambio grande en el patrón de ésta.

Figura 13. Ejemplo del tercer filtro aplicado, Logaritmo en base 10.



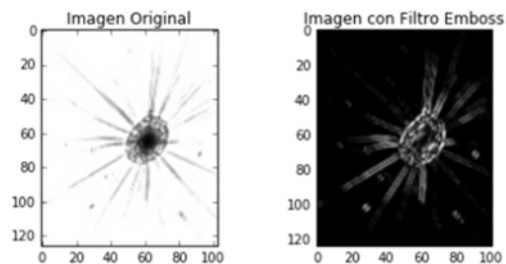
Se aplicó un filtro sharpening a las imágenes para realzar los bordes, que son características de alto nivel detectadas y aprendidas por los filtros convolucionales. Al mejorar los bordes también mejora en general la calidad de la imagen, se resaltan detalles y se tiene más clara la forma del organismo.

Figura 14. Ejemplo del cuarto filtro aplicado, Sharpen.



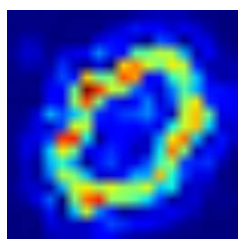
Utilizando un filtro emboss, que también resalta los bordes, dando un efecto tridimensional la imagen, que queda en escala de grises y los bordes color negro, bastante resaltados.

Figura 15. Ejemplo del quinto filtro aplicado, filtro Emboss.



Para finalizar, se aplicó un filtro de convolución en las imágenes.

Figura 16. Ejemplo del sexto filtro aplicado, Filtro de convolución.



6.2 FASE 2 DISEÑO Y DESARROLLO

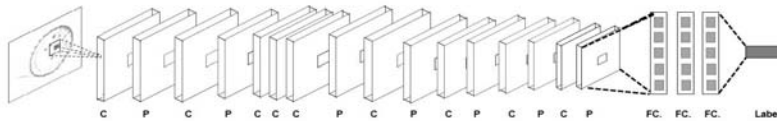
Se realizaron varios diseños antes de llegar a los listados a continuación. En total fueron diseñadas 16 arquitecturas de CNN. Se describe la primera red diseñada y posteriormente se encuentran descritas las 4 arquitecturas con mejor desempeño. La fase de validación fue realizada con el 25% de las imágenes provistas para el entrenamiento, elegidas de forma aleatoria en cada caso.

6.2.1 Diseño de la red: Antes de comenzar a diseñar las redes, se estudiaron las redes más conocidas y con mejores desempeños de clasificación de objetos en general. Se revisó la arquitectura y el desarrollo de las redes AlexNet, VGG, GoogLeNet, ZFNet, en general se encontró que cuentan con grupos de entrada de datos relativamente pequeños que no superan los 20 y en general son redes no muy grandes entre 5 y 13 capas de convolución lineales, exceptuando GoogLeNet, que es una red con gran complejidad conformada por varias capas de inyección a lo ancho y profundo de la misma. Éstas cuentan con 2 ó 3 capas de producto interno, capas de pooling con filtro 3. Los parámetros en las capas convolucionales fueron bastante variables en las redes.

Viendo la similitud de dichos diseños, se diseñó una red iniciando con la capa de entrada, dividiendo los datos en grupos de 20, buscando no saturar la red. Una capa convolucional con filtro tamaño 5 seguido de una capa ReLu y una de normalización. En total la red está conformada por 9 capas convolucionales, seguidas por su respectiva capa ReLu para rectificar las características aprendidas y obtener reconstrucciones válidas de éstas en cada capa. Buscando disminuir la dimensionalidad de los datos a procesar por la red, se añadieron 7

capas de pooling. Finalizando con 3 capas de producto interno, para calcular las funciones no lineales de las características aprendidas en las capas anteriores.

Figura 17. Arquitectura del primer diseño de red realizado.

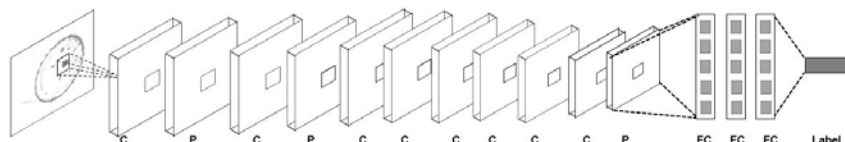


6.2.2 Desarrollo del prototipo: Éste primer prototipo fue desarrollado en la plataforma de Nvidia, Digits con código base PyCaffe, lamentablemente no se obtuvo el resultado esperado con un desempeño del 6.56% de acierto en la clasificación de las imágenes.

Es muy probable que estos resultados se deban a la gran cantidad de capas de pooling en la red, disminuyendo en exceso a cantidad de datos y generando una pérdida significativa de la información.

6.2.3 Diseño de la red: Esta red cuenta con una capa de entrada de datos, que recibe las imágenes de 256 x 256, dividiéndolas en grupos de 50 que se trabajarán simultáneamente por la red durante el entrenamiento. Seguida por una capa de convolución con un filtro tamaño 11 y paso 4 que busca extraer algunos rasgos de las imágenes como bordes y curvas. En el siguiente nivel se tiene una capa ReLu que realiza una función de activación sobre los datos y una capa de normalización. Luego se encuentra una capa de pooling con filtro 3 para disminuir la dimensión de los datos de salida, seguida por otra capa convolucional con filtro 5, paso 1 y un padding 2, que aprenderá nuevas características de los datos. Su salida será normalizada y pasará por una capa de función de activación (ReLu) entrará a una capa de pooling de filtro 3. Estas capas son seguidas de 3 capas de convolución de filtro 3 y paso 1, con su respectiva capa ReLu, una con filtro 6 seguida de 2 capas convolucionales de filtro 3. Estos resultados ingresan a una capa de pooling, que hará una última disminución de la dimensionalidad antes de ingresar estos datos a 2 capas completamente conectadas que calcularán las funciones no lineales de las características aprendidas por las capas convolucionales. Se finaliza con las capas de pérdida Softmax y de desempeño.

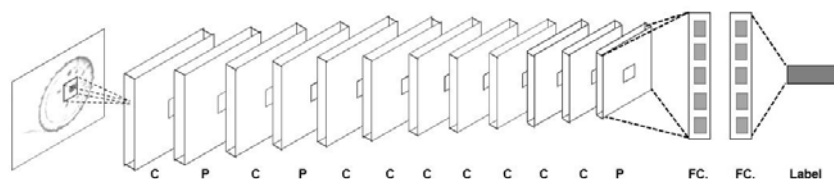
Figura 18. Arquitectura del segundo diseño de red.



6.2.4 Desarrollo del prototipo: Se creó una red como se estableció en el diseño, por medio del framework Digits, agregando padding en algunas de las capas convolucionales. El entrenamiento y el diseño tuvieron un efecto deseado con un acierto de 81.20% gracias sobre las 60,672 imágenes separadas para la validación, correspondientes al 25% de las imágenes de entrenamiento provistas después de obtener las copias rotadas sobre su eje cada 45°, separando este 25% de cada una de las clases existentes al azar, teniendo cuidado de no utilizar para la validación una rotación de una de las imágenes vistas durante el entrenamiento, por tanto se eligió la imagen con todas sus rotaciones para la carpeta de validación. Es posible que para este tipo de datos sea mejor trabajar con 2 capas completamente conectadas como en la red VGGNet.

6.2.5 Diseño de la red: Viendo la mejora en el desempeño y se decidió crear una nueva red similar a la anterior, agregando una capa de convolución adicional antes de la última capa de pooling con filtro 3 y paso 1, teniendo en cuenta que en éstas se extraen las características de alto nivel. Se eliminó una de las capas completamente conectadas, quedando sólo con 2 de éstas antes de pasar a las capas de pérdida y acierto.

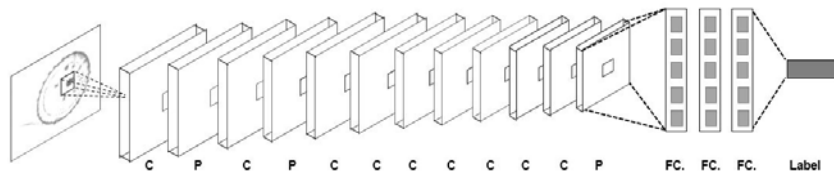
Figura 19. Arquitectura del tercer diseño de red.



6.2.6 Desarrollo del prototipo: Se agregó también un padding de 1 a la nueva capa convolucional, pero no fue ubicada como la última capa de éste tipo de la red, sino como la penúltima, para no modificar el tamaño de la salida de ésta. A pesar de haber tenido un buen porcentaje de acierto con 80.76% su desempeño no mejoró al de la red anterior.

6.2.7 Diseño de la red: Al eliminar una capa completamente conectada y agregar una capa convolucional disminuyó un poco el porcentaje de acierto, aunque teniendo en cuenta que las imágenes tomadas para cada entrenamiento son diferentes existe un margen de error en la clasificación, por lo que es posible decir que tiende a estabilizarse, sin embargo se procedió a tomar el diseño anterior y agregar nuevamente la capa de producto interno manteniendo todos los parámetros previamente definidos. Se miró también la posibilidad de mejoría si se aumentaba el tamaño de grupo de entrada de 50 a 60 imágenes.

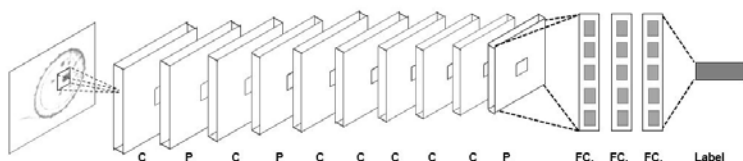
Figura 20. Arquitectura del cuarto diseño de red.



6.2.8 Desarrollo del prototipo: El modelo fue implementado de la forma descrita, sin embargo el desempeño de la red cayó levemente, con un porcentaje de acierto final del 80.51%. Por tanto los grupos de datos de entrada no pueden ser tan grandes.

6.2.9 Diseño de la red: En un último intento de mejora del porcentaje de acierto, se decidió crear una red con menos capas convolucionales y modificar algunos parámetros. De los datos de entrada se crearon grupos de 60 imágenes por proceso. A la primera capa de convolución le fue aumentado el tamaño del filtro a 12 y disminuido el paso a 3. La siguiente capa modificada fue la segunda convolucional donde se aumentó en uno el tamaño del filtro y del paso.

Figura 21. Arquitectura del quinto diseño de red.



6.2.10 Desarrollo del prototipo: En este momento se decidió aumentar el tamaño de la salida de la cuarta capa convolucional y aumentar a 2 el padding de la tercera. Obteniendo con menos capas un resultado muy similar a los anteriores, un desempeño de 80.71% de acierto en la clasificación de imágenes nuevas para el modelo.

6.3 FASE DE ENTRENAMIENTO

Se tomaron los 2 filtros con mejor porcentaje de clasificación en general y fueron rotados para proporcionar a la red nuevos datos, dado que la literatura sugiere que esta forma de entrenamiento puede aumentar el desempeño de la red.

6.3.1 Primera variación: Inicialmente se tomó el primer procesado de imágenes y se aplicó una variación en las rotaciones de las imágenes sobre el eje. Para las clases con mayor cantidad de datos, fueron rotadas cada 30° y para las clases con menor cantidad de datos suministrados, se realizó una rotación a la imagen cada 20° .

6.3.2 Segunda rotación: Se tomaron las imágenes con filtro Sharpen y fueron rotadas cada 45° añadiendo también un poco de ruido en ellas y se procedió a entrenarlas con todas las redes diseñadas.

7. RESULTADOS OBTENIDOS

A continuación se encuentra una tabla con los resultados obtenidos para cada red diseñada y para cada uno de los procesamientos realizados, en orden descendente y de izquierda a derecha. Al reportar un porcentaje tan bajo no se entrenó la primera red diseñada con el resto de grupos de imágenes.

P1: Primer procesado.

NP: Imágenes originales.

P2: Rotaciones dependiendo de la cantidad de datos por clase.

P3: Transformada rápida de Fourier.

P4: Logaritmo en base 10.

P5: Filtro Sharpen.

P6: Filtro Emboss.

P7: Convolución.

P8: Rotación cada 45° a imágenes con filtro Sharpen

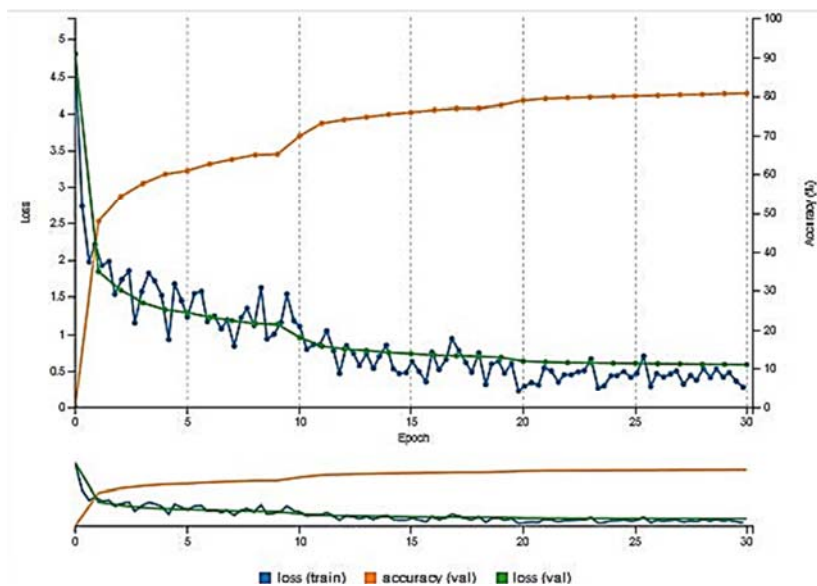
Tabla 2. Resultados Obtenidos.

RED	P1	NP	P2	P3	P4	P5	P6	P7	P8
Diseño 2	81.20	70.43	75.64	60.07	70.00	70.43	69.33	62.89	71.25
Diseño 3	80.76	69.63	74.30	59.66	68.24	69.77	68.33	61.95	71.59
Diseño 4	80.51	68.86	74.92	57.76	68.54	68.61	68.16	61.67	69.55
Diseño 5	80.71	71.09	74.71	60.63	70.24	70.79	70.58	62.98	71.24

Se puede observar que los mejores resultados se dieron para el primer procesado de las imágenes y no está tan relacionado con el hecho de generar más datos de entrenamiento ficticios, pues esto también se hace para el último procesado que se muestra y los resultados están bastante lejos de los obtenidos con éste primer procesamiento de las imágenes porque no solo realza los bordes sino que resalta en general las características del objeto deseado y elimina adicionales que son fuente de error.

La red con mejor desempeño cuenta con 8 capas convolucionales y 3 capas de pooling con un filtro, tamaño de paso y de grupo de datos relativamente grande en comparación a lo visto en la literatura, pero fue necesario por la poca variación en características de los objetos en clases, no son fácilmente distinguibles y necesita extraer muchos detalles minuciosos en cada imagen. Esta red inmediatamente aprende características y puede clasificar con un gran porcentaje de acierto, como se puede ver en la figura 13. El aprendizaje se realiza en su gran mayoría hasta la décima iteración, desde este punto el aumento en el porcentaje de acierto aumenta muy poco con cada iteración, mostrando que extrae algunas nuevas características de las clases, pero no son tan significativas y se vuelve casi nulo a partir de la iteración 20.

Figura 22. Porcentaje de acierto en la red con mejor desempeño a lo largo de las 30 iteraciones.



Los entrenamientos de los 4 primeros diseños de redes fueron realizados en GPUs Tesla M2075 al 99% y duraron en promedio 2 días y medio en el framework Digits para los datasets con mayor cantidad de imágenes y 10 horas para los

datasets con la cantidad original, más las 2 horas requeridas para el procesamiento de las imágenes en python.

La red con mejor desempeño duró 2 días y 11 horas y el procesamiento de las imágenes con rotación tardó 5 horas, para un total 2 días y horas, siendo la red con más costo computacional.

El último diseño al tener menos complejidad, tuvo menos costo computacional, tardando 4 horas para los dataset con menor cantidad de datos (imágenes sin rotaciones) y alrededor de 10 horas para los dataset con rotaciones de las imágenes

En su primera capa de convolución, donde extra la mayor cantidad de características primitivas, es posible obtener una imagen con su representación, que sale de una media generada por el framework de todas las imágenes de dataset. En la figura 23, es posible ver las características extraídas por la primera capa de convolución y en la figura 24, la imagen media generada de los datos de entrenamiento.

En la imagen de las características extraídas, es posible ver que dicha red pudo encontrar patrones de bordes y degradado en el color de los datos de forma suficientemente clara para aprender las características de cada clase.

Figura 23. Características extraídas en la primera capa de convolución.

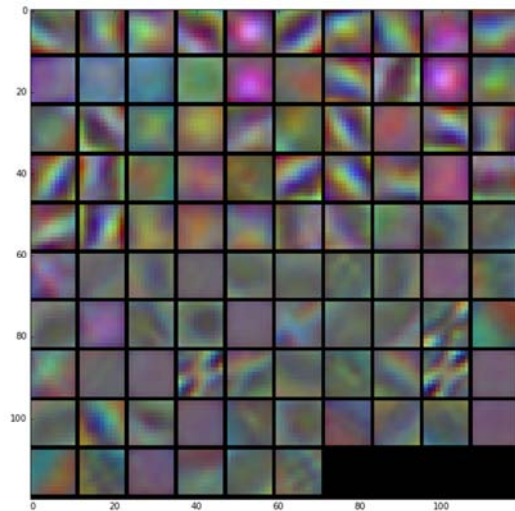
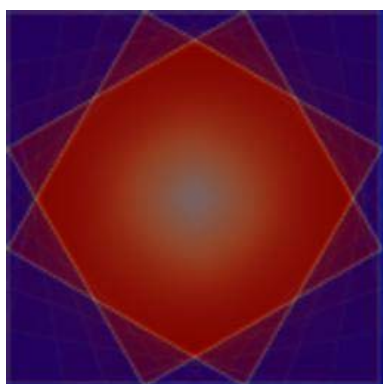


Figura 24. Media de las imágenes con el primer procesado realizado.



8. CONCLUSIONES

- Un adecuado procesamiento de las imágenes, previo al entrenamiento representa un gran porcentaje del desempeño de clasificación, con una diferencia de casi 20% con un buen procesamiento y uno más sencillo.
- Los parámetros de la primera capa convolucional y tamaño de datos de entrada afectan el desempeño de la red.
- Cambiar el modo de entrenamiento al añadir nuevas imágenes con variaciones y perturbaciones no proporcionó ninguna mejora en el desempeño, por el contrario su efecto fue negativo.
- A pesar que las características de alto nivel de las imágenes son aprendidas en las capas de convolución, no hay necesariamente un mejor desempeño al tener mayor cantidad de éstas en la red.
- Agregar muchas capas de pooling o agrupamiento afecta negativamente el desempeño de las CNN al disminuir demasiado la dimensionalidad de los datos, dejando muy poca información.
- El parámetro más importante en las capas de convolución de las CNN es el tamaño del filtro, más que el tamaño del paso, si es muy pequeño hay pérdida en la capacidad de aprendizaje, si es muy grande, las características aprendidas son muy generalizadas y el desempeño también será bajo.
- Se probaron algunos diseños de red con una aparente gran complejidad y gran cantidad de capas, que reportaron resultados tan bajos que no se consideraron para la presentación de éste trabajo. Pero contrario a lo que instintivamente se creía, mayor cantidad de capas y complejidad no significa mejor desempeño.
- El FlowCybot, que captura imágenes en vivo y las clasifica tiene un porcentaje de acierto del 88% pero sólo funciona en las 22 especies de fitoplancton, mientras que por medio de la mejor red diseñada, se obtuvo un 81.2032% para las 121 clases de plancton definidas por el centro marítimo de la Universidad de Óregon. FlowCybot fue entrenado con 6,600 imágenes

Comentado [r1]: En las conclusiones se tiene que ver explícitamente que has abordado cada objetivo específico planteado. Pon al menos un punto que se vea que responde DIRECTAMENTE a cada uno de los objetivos.

pertenecientes a 22 categorías distribuidas uniformemente, con 300 imágenes por clase. En ambos casos se trabajaron imágenes de resolución variable.

- Tienen un alto coste computacional debido a la gran cantidad de datos usados para su entrenamiento, pero teniendo en cuenta el buen desempeño logrado y comparándolo con el tiempo que demora el proceso de clasificación manual, se le considera un costo justo.

9. REFERENCIAS BIBLIOGRÁFICAS

- [1] BISHOP, C. M. «Pattern Recognition and Machine Learning». Springer. 2006.
- [2] DAVIS, C.S «et al. The Video Plankton Recorder (VPR) Design and Initial Results» 1992
- [3] DEEP LEARNING, Convolutional Neural Networks (LeNet) Tutorial [online] Available from the Internet <http://deeplearning.net/tutorial/lenet.html>
- [4] KARPATHY, Andrej. «CS231n Convolutional Neural Networks for Visual Recognition» [online] 2015 Available from internet <<http://cs231n.github.io/convolutional-networks>>
- [5] KOREKADO, Keisuke et al. «A Convolutional Neural Network VLSI for Image Recognition Using Merged/Mixed Analog-Digital Architecture». 2003.
- [6] LARRAÑAGA, P. et al. «Redes Neuronales. Universidad del País Vasco – Euskal Herriki Unibertsitatea.
- [7] NIELSEN, Michael. «Neural Networks and Deep Learning». Determination Press, 2015.
- [8] SOSIK, Heidi et al. «Automated taxonomic classification of phytoplankton sampled with imaging-in-flow cytometry». American Society of Limnology and Oceanography, Inc, 2007.
- [9] XIAOOU Tang et al. «Automatic Plankton Image Recognition» 1998
- [10] ZEILER, Matthew. «Stochastic Pooling for Regularization of Deep Convolutional Neural Networks». Cornell University Library, 2013.
- [11] <https://www.igd.fraunhofer.de/en/Institut/Abteilungen/MAG/Projekte/ZooCount-Semi-automatic-segmentation-and-classification>.
- [12] http://www.teachoceanscience.net/teaching_resources/education_modules/plankton_-_aquatic_drifters/

BIBLIOGRAFÍA

CAFFE DOCUMENTATION [online]

⟨<http://christopher5106.github.io/deep/learning/2015/09/04/Deep-learning-tutorial-on-Caffe-Technology.html>⟩

CAFFE DOCUMENTATION FOR DEEP LEARNING [online]

⟨https://docs.google.com/presentation/d/1UeKXVgRvvxg9OUdh_UiC5G71UMscNPvArsWER41PsU/edit#slide=id.gc2fcdcce7_216_0⟩

LECUN, Yann. Convolutional Neural Networks for Images, Speech and Time-Series. Université de Montréal. 2013. [online]

⟨https://www.researchgate.net/profile/Yann_Lecun/publication/2453996_Convolutional_Networks_for_Images_Speech_and_Time-Series/links/0deec519dfa2325502000000.pdf⟩

LIU, Tianyi et al. Implementation of Training Convolutional Neural Networks. University of Chinese Academy of Sciences, Beijing, China. 2015 [online]

⟨<https://arxiv.org/ftp/arxiv/papers/1506/1506.01195.pdf>⟩

SCIKIT-LEARN DOCUMENTATION [online] ⟨<http://scikit-learn.org/stable/documentation>⟩

SZELISKI, Richard. Computer Vision: Algorithms and Applications, Electronic Draft. [online] ⟨<http://szeliski.org/Book/>⟩