

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

Visión Electrónica

Más que un estado sólido

<https://revistas.udistrital.edu.co/index.php/visele>

Visión Electrónica

VISIÓN DE CASO

Sistema automático de clasificación de peces

Automatic fish classification system

Robinson Jiménez Moreno¹, Javier Eduardo Martínez Baquero², Luis Alfredo Rodríguez Umaña³

INFORMACIÓN DEL ARTÍCULO

Historia del artículo:

Enviado: 21/09/2018

Recibido: 30/10/2018

Aceptado: 04/12/2018

Palabras clave:

Aprendizaje de máquina

Aprendizaje profundo

Reconocimiento de patrones

Clasificación de peces

Visión de máquina

RNC

Open access



Keywords:

Machine learning

Deep learning

Pattern recognition

Fish classification

Machine vision

CNN

RESUMEN

El presente artículo expone el diseño de una arquitectura de red para reconocimiento de patrones orientada a la clasificación automática de dos tipos de peces: mojarra y tilapia. Se emplea una arquitectura basada en aprendizaje profundo mediante una red neuronal convolucional (RNC) para la cual se determina la base de datos a emplear y los diferentes hiperparámetros que la componen. Se logra obtener, mediante análisis por matriz de confusión, un desempeño del 100% de la red bajo las condiciones controladas el sistema de clasificación, es decir: color de banda transportadora uniforme y uso de luz día.

ABSTRACT

The present article exposes the design of a network architecture for pattern recognition, oriented to the automatic classification of two types of fish: mojarra and tilapia. An architecture based on deep learning is used by means of a convolutional neuronal network (CNN), for which the database to be used and the different hyperparameters that compose it are determined. It is possible to obtain, through confusion matrix analysis, a 100% performance of the network under the controlled conditions of the classification system, that is: uniform conveyor belt color and daylight use.

¹ Ing. Electrónico, Universidad Distrital Francisco José de Caldas, Colombia. MSc. en Automatización Industrial, Universidad Nacional de Colombia, Afilación institucional: Docente de planta Universidad Militar Nueva Granada, Colombia. Correo electrónico: robinson.jimenez@unimilitar.edu.co. ORCID: <https://orcid.org/0000-0002-4812-3734>

² Ing. Electrónico, Universidad de los Llanos, Colombia. Esp. en Instrumentación Electrónica, Universidad Santo Tomás, Colombia. Msc en Tecnología Educativa y medios innovadores para la Educación, Universidad Autónoma de Bucaramanga, Colombia. Afilación institucional: Docente de Planta Universidad de los Llanos, Colombia. Correo Electrónico: jmartinez@unillanos.edu.co. ORCID: <https://orcid.org/0000-0003-4377-7867>

³ Ing. Electrónico, Universidad de los Llanos, Colombia. Esp. en Automática e Informática Industrial, Universidad Autónoma de Colombia, Colombia. Afilación institucional: Docente de Planta Universidad de los Llanos, Colombia. Correo electrónico: lrodriguez@unillanos.edu.co. ORCID: <https://orcid.org/0000-0001-7346-5640>

1. Introducción.

La clasificación de imágenes se ha convertido en un proceso cada vez más importante de investigación aplicada en la industria y el comercio. Una prueba de ello es la conocida competencia de clasificación de imágenes a gran escala: ILSVRC, [1]. El ILSVRC usa 1,2 millones de imágenes de 1.000 categorías de la base de datos de ImageNet, [2].

De otra parte, en general las CNN son un tipo de redes neuronales que se han convertido en un algoritmo popular para resolver la clasificación de imágenes luego de que AlexNet ganara el ILSVRC de 2012, [3]. Desde entonces se han creado innumerables CNN, e ILSVRC se estableció como un test de referencia común para estas redes. En este sentido, las CNN se han aplicado para automatizar tareas tales como el reconocimiento de documentos [4], el reconocimiento de números y letras, la detección de rostros [5-6], el reconocimiento de objetos [7], la clasificación de imágenes [8], la estimación de visibilidad de imágenes [9], y al aprendizaje de sonidos ambientales, [10].

De otro lado, las aplicaciones que involucran sistemas de visión de máquina para reconocimiento de peces tienen múltiples aplicaciones y llevan varios años en desarrollo. En [11] se expone un sistema interactivo de seguimiento para peces en un acuario para enseñanza de los visitantes. Estos sistemas interactivos permiten el aprendizaje tanto de máquinas como de humanos; de ahí que en [12] se desarrolle un sistema de entrenamiento, reconocimiento y consulta de peces basado en imágenes digitales.

En particular, en el reconocimiento de peces como objeto de investigación se han tenido múltiples antecedentes. En [13] se presenta el uso de un vector de cuantización por histograma con 94% de detección para un mismo tipo de pez. Tales aplicaciones han incidido significativamente la industria de peces. Es así como en [14] se ha empleado un sistema de reconocimiento de imagen, basado en color, para el reconocimiento de peces enfermos, lo cual permitió su posterior clasificación.

El estado del arte presenta, incluso, aplicaciones bioinspiradas en el nado del pez en cardumen para construir algoritmos de reconocimiento de patrones en imágenes, [15]. Sin embargo, la tarea fundamental en la industria es inversa, es decir: emplear algoritmos de reconocimiento de patrones en

imágenes para identificación de peces, por ejemplo, para conteo de los mismos, como se presenta en [16].

En términos de la presente investigación, las técnicas más recientes empleadas para las tareas descritas anteriormente son las CNN, como se expone [17]. Con esta técnica se han desarrollado aplicaciones de reconocimiento de peces orientadas a aplicaciones de sostenibilidad ambiental, demostrando un alto desempeño en dicho reconocimiento.

En consecuencia, la técnica basada en CNN se emplea en la investigación descrita por el presente artículo para la clasificación de dos especies de peces, tilapia y mojarra, en una línea de producción. Aquí, entonces, se expone el diseño y los resultados de una CNN entrenada para este propósito, obteniéndose -mediante una arquitectura robusta- un desempeño del 100% de clasificación de las dos especies en un ambiente controlado. Este tipo de aplicaciones tiene impacto a nivel industrial, derivado del proceso de automatización que implica, lo cual permite reducir los tiempos de clasificación manual, así como un trabajo continuo.

El documento se estructura en tres secciones, la primera corresponde a los métodos y materiales empleados; la segunda sección corresponde al análisis de resultados; y la tercera sección a las conclusiones obtenidas.

2. Métodos y materiales

El sistema de clasificación diseñado se orienta al esquema planteado en la Figura 1. Se emplea un ambiente controlado basado en iluminación de luz día, una cámara fija de vista superior y una banda transportadora sobre la cual ira el pescado. La banda con fondo blanco y delimitadores laterales negros, transportará el pescado desde el punto de descarga de este hasta el punto de clasificación. En el Intermedio del trayecto se ubica la cámara web para captura de la escena, que a su vez envía la imagen a un computador sobre el cual corre el algoritmo de clasificación, [18].

El primer paso consiste en establecer la base de datos con la cual el sistema de clasificación es entrenado; para el caso se emplean dos clases de pescado: tilapia y mojarra. Se utilizan 100 imágenes de cada clase en las diferentes ubicaciones en que pueden aparecer en la imagen; es decir: si bien el sistema solo permite el paso de un pescado a la vez, en su orientación respecto a la cámara puede suceder

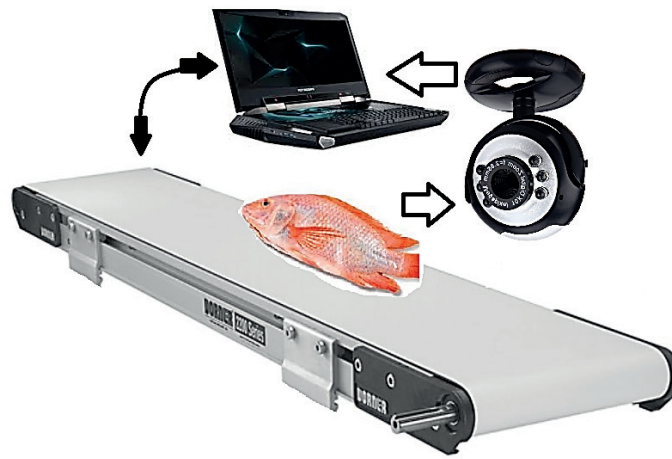
que esté mirando hacia el frente o hacia atrás, con la aleta dorsal hacia el borde lateral izquierdo o lateral derecho de la banda. En algunos casos es posible que se aprecien bordes de dicha banda en la imagen, dichas variaciones se ilustran en la Figura 2.

De la base de datos se emplea el 80% para entrenamiento, y el 20% para validación de la clasificación. Estas son adquiridas mediante una cámara web convencional con resolución de imagen

de 640 X 480 píxeles, las cuales son redimensionadas al interior del algoritmo a 180 X 180 píxeles a fin de reducir el costo computacional del entrenamiento, esta dimensión es la máxima utilizable por equipo empleado, el cual se describe más adelante.

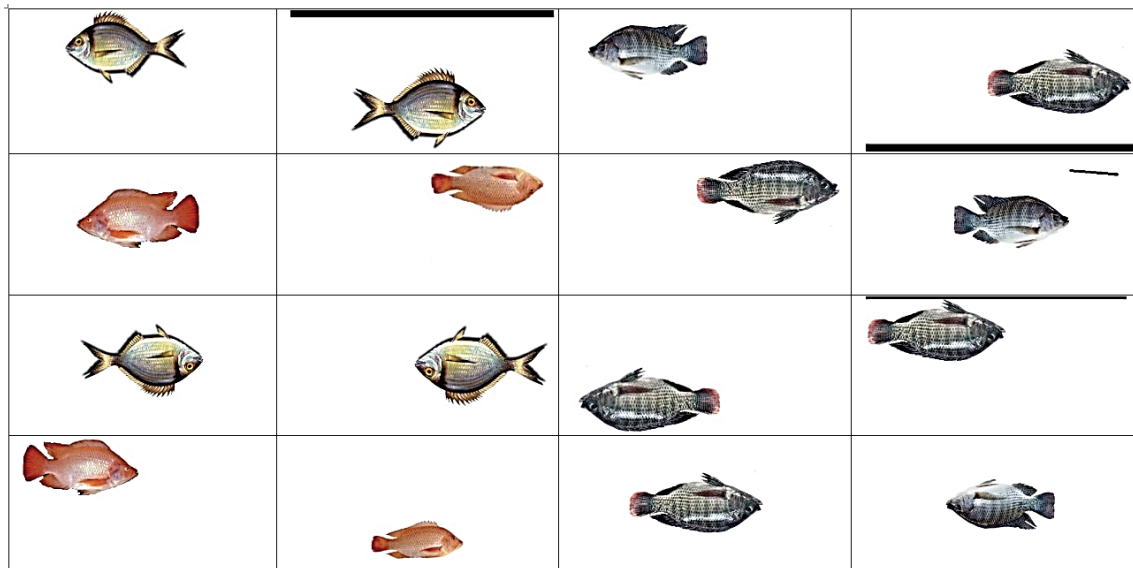
Para determinar la arquitectura de red adecuada se deben variar los diferentes parámetros de la misma, donde una CNN se compone de dos etapas: la de extracción de características, y la de clasificación.

Figura 1. Esquema de clasificación.



Fuente: elaboración propia.

Figura 2. Base de datos de entrenamiento



Fuente: elaboración propia.

En la etapa de extracción se deben determinar cuántas capas convolucionales tendrá, y cada una de estas cuántos filtros empleará.

El cálculo de los parámetros se establece en función del volumen de entrada y salida de cada capa de convolución mediante las ecuaciones (1) a (3); donde W_{n+1} y H_{n+1} corresponden al volumen de salida (Ancho y alto respectivamente, D_{n+1} representa la profundidad), W_n y H_n corresponden al volumen de entrada, F al tamaño lateral del filtro de convolución, el cual es cuadrado, S corresponde al parámetro de desplazamiento del filtro sobre el volumen de entrada, y P al relleno lateral requerido en la operación de convolución en las fronteras de la imagen, [19].

$$W_{n+1} = \frac{W_n - F + 2P}{S} + 1 \tag{1}$$

$$H_{n+1} = \frac{H_n - F + 2P}{S} + 1 \tag{2}$$

$$D_{n+1} = K_n \tag{3}$$

Para el caso se emplea la arquitectura de red ilustrada en la Tabla 1, la cual consta de 3 capas de convolución en la etapa de extracción de características, y dos *fully connected* con un 50% de *dropout* en la etapa

de clasificación, y una *fully connected* de salida final. La columna de *kernel* presenta la cantidad de filtros empleados y el tamaño de cada uno.

Tabla 1. Arquitectura red empleada

LAYER	KERNEL	
Input	180x180x3	
Convolution/RELU	98	8
Convolution/RELU	192	6
MaxPooling		3
Convolution/RELU	192	3
MaxPooling	3	
Convolution/RELU	320	3
MaxPooling	2	

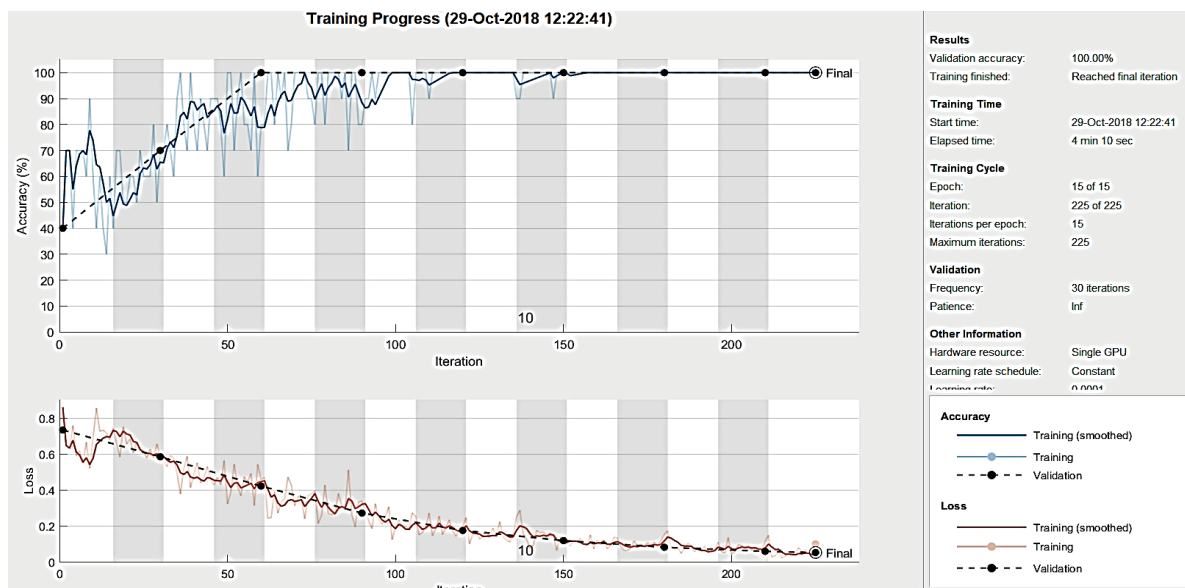
Fuente: elaboración propia.

3. Análisis de resultados

La Figura 3 ilustra el desempeño obtenido en el entrenamiento de la red. Para el caso se observa que llega al 100% de precisión en la clasificación después de 225 iteraciones, tomando poco más de 4 minutos en un equipo de cómputo con GPU de 8 GB de memoria, y una tarjeta NVIDIA 1050.

La Figura 4 permite observar algunas de las activaciones de la primera capa de convolución, donde se evidencia el aprendizaje de los filtros. Se

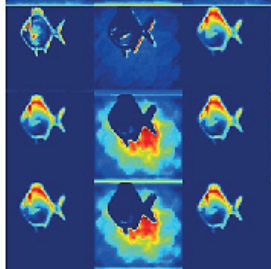
Figura 3. Desempeño del entrenamiento



Fuente: elaboración propia.

nota claramente cómo identifica el pez (columnas laterales), y el fondo (columna central).

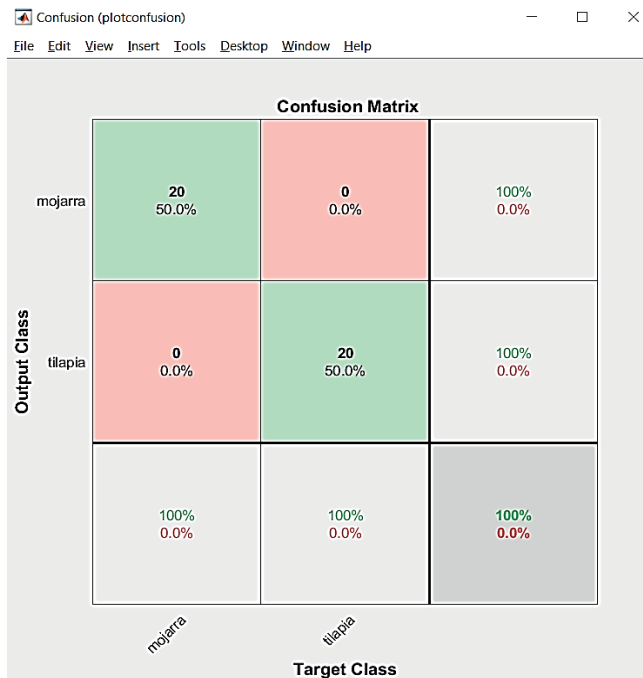
Figura 4. Activaciones de aprendizaje



Fuente: elaboración propia.

La matriz de confusión de la Figura 5 ilustra el desempeño en la clasificación con los datos de validación, evidenciando la correcta identificación de cada una de las dos clases de pez empleadas. Esto indica que no hay errores en el desempeño de la red, teniendo en cuenta que las condiciones de iluminación empleada no deben variar considerablemente. Para el caso se obtuvieron tiempos de clasificación promedio de 0,4 segundos, lo que permite el uso del algoritmo en aplicaciones de visión de máquina, como la propuesta, en tiempo real.

Figura 5. Matriz de confusión



Fuente: elaboración propia.

4. Conclusiones

Los sistemas de visión de máquina son un complemento importante en los sistemas de automatización, los cuales emplean técnicas de reconocimiento de patrones para operar. Dentro de estas técnicas, CNN demostraron una alta eficiencia en el reconocimiento de las dos clases de peces establecidas, con tiempos de clasificación reducidos respecto a las velocidades típicas de operación de las bandas transportadoras usadas en la industria de alimentos.

La red obtenida permite diferenciar claramente el fondo del elemento que transporta, pero se deben considerar variaciones de dicho fondo, que con el transcurso del tiempo pueden cambiar la tonalidad, debido a que al ensuciarse por el obvio uso puede reducir el desempeño de clasificación obtenido.

En perspectiva de trabajos de investigación futuros se tienen experiencias donde se han usado máquinas de resonancia magnética nuclear C (C NMR) en combinación con una red neuronal probabilística (PNN), [20]. La máquina C NMR se utiliza para determinar estructuras moleculares y compuestos individuales, y para examinar la cinética de reacciones específicas se utiliza magnetismo [21]. Los experimentos alcanzan precisiones de 98.5% con PNN. Los peces utilizados proceden de Noruega, Escocia, Canadá, Islandia, Irlanda, las Islas Feroe y Tasmania, y la clasificación de los orígenes ha alcanzado el 82,2% de precisión con PNN. Esta experiencia también utiliza tejido muscular de los peces que no se ajustan bien a un ecosistema de peces vivos.

En otra línea de investigación, ha habido un incremento de soluciones usando inteligencia artificial para clasificar diferentes tipos de peces [22] [23] [24]. Todos estos utilizan el conjunto de datos del proyecto Fish4Knowledge [25], que consta de 27.370 imágenes de peces exóticos categorizados en 23 especies.

En el anterior sentido, la CNN -Fast R-CNN- es uno de los modelos probados con el conjunto de datos de Fish4Knowledge [22]. Los experimentos realizados en el artículo utilizan un subconjunto de Fish4Knowledge, limitando el número de clases a aquellos con más de 600 imágenes. La detección de objetos se utiliza para extraer solo los peces de las imágenes. El enfoque adoptado comienza con el entrenamiento previo de una AlexNet [18] en la base

de datos de ImageNet. Luego se modifica la AlexNet para entrenar en el subconjunto de Fish4Knowledge y aplicar los pesos a Fast R-CNN. La Fast R-CNN toma como ponderación las propuestas de ponderaciones y regiones preentrenadas hechas por AlexNet, alcanzándose una precisión promedio de 81.4%.

La propuesta dada en [24], entrena previamente a un CNN similar a AlexNet con cinco capas convolucionales y tres capas completamente conectadas, utilizando 1.000 imágenes de 1.000 categorías del conjunto de datos de ImageNet, y aplica los pesos al conjunto de datos de Fish4Knowledge, utilizando solo 50 imágenes por categoría y 10 categorías del conjunto de datos de Fish4Knowledge durante la formación. Las imágenes del conjunto de datos de Fish4Knowledge se preprocesan utilizando eliminaciones. La precisión alcanzada en las 1.420 imágenes de prueba es 85.08% utilizando cantidades muy pequeñas de datos.

La precisión más alta lograda en este conjunto de datos es 98.64% usando una CNN con una función clasificadora después de aplicar filtros a las imágenes originales para extraer la forma del pez y retire el fondo [23]. Esta red se llama DeepFish, y consiste en tres capas convolucionales estándar y tres capas totalmente conectadas.

Es decir, la solución propuesta en el presente artículo es plausible; y a gran escala, para el medio colombiano, requerirá complementarla con Hardware y CNN de mayor robustez.

5. Reconocimientos

Los autores reconocen a los grupos de investigación DAVINCI, de la Universidad Militar Nueva Granada; y GITECX, de la Universidad de los Llanos, entre los cuales se dio la sinergia para el desarrollo conjunto de esta investigación.

Referencias

- [1] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg y L. Fei-Fei, “ImageNet Large Scale Visual Recognition Challenge” *International Journal of Computer Vision (IJCV)*, vol. 115, n° 3, pp. 211–252, 2015, <https://doi.org/10.1007/s11263-015-0816-y>
- [2] J. Deng, W. Dong, R. Socher, L. Li, K. Li y L. FeiFei, “Imagenet: A large-scale hierarchical image database” In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 248–255. IEEE, 2009.
- [3] ImageNet, “Imagenet large scale visual recognition challenge 2012 (ilsvrc2012)” [En línea], Disponible en: <http://www.image-net.org/challenges/LSVRC/2012/>
- [4] P. Y. Simard, D. Steinkraus, J.C. Platt, et al, “Best practices for convolutional neural networks applied to visual document analysis” In *ICDAR*, vol. 3, pp. 958–962, 2003, <https://doi.org/10.1109/ICDAR.2003.1227801>
- [5] F. Nasse, C. Thureau y G.A. Fink, “Face detection using GPU-based convolutional neural networks” In *International Conference on Computer Analysis of Images and Patterns*, pp. 83–90. Springer, 2009, https://doi.org/10.1007/978-3-642-03767-2_10
- [6] Y. Taigman, M. Yang, M. Ranzato y L. Wolf, “Deepface: Closing the gap to human-level performance in face verification” In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1701–1708, 2014, <https://doi.org/10.1109/CVPR.2014.220>
- [7] S. Ren, K. He, R. Girshick y J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks” In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems 28*, pp. 91–99. Curran Associates, Inc., 2015.
- [8] T. Guo, J. Dong, H. Li y Y. Gao, “Simple convolutional neural network on image classification,” 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA) (, Beijing, 2017, pp. 721-724, <https://doi.org/10.1109/ICBDA.2017.8078730>
- [9] A. Giyenko, A. Palvanov y Y. Cho, “Application of convolutional neural networks for visibility estimation of CCTV images,” 2018 International Conference on Information Networking (ICOIN), Chiang Mai, 2018, pp. 875-879, <https://doi.org/10.1109/ICOIN.2018.8343247>

- [10] B. Zhu et al., “Learning Environmental Sounds with Multi-scale Convolutional Neural Network,” 2018 International Joint Conference on Neural Networks (IJCNN), Rio de Janeiro, 2018, pp. 1-8, <https://doi.org/10.1109/IJCNN.2018.8489641>
- [11] Jia-Hong Lee, Mei-Yi Wu y Zhi-Cheng Guo, “A tank fish recognition and tracking system using computer vision techniques,” *2010 3rd International Conference on Computer Science and Information Technology*, Chengdu, 2010, pp. 528-532, <https://doi.org/10.1109/ICCSIT.2010.5563625>
- [12] X. Zheng, J. Zhong y Y. Zhang, “A SOA-Based Fish Recognition System Prototype,” *2011 Third International Conference on Intelligent Human-Machine Systems and Cybernetics, Zhejiang*, 2011, pp. 66-69, <https://doi.org/10.1109/IHMSC.2011.86>
- [13] Y. Nishida, T. Ura, T. Hamatsu, K. Nagahashi, S. Inaba and T. Nakatani, “Fish recognition method using vector quantization histogram for investigation of fishery resources,” *2014 Oceans–St. John’s*, St. John’s, NL, 2014, pp. 1-5, <https://doi.org/10.1109/OCEANS.2014.7003268>
- [14] Y. Wang, H. Ye y B. Li, “A research based on recognition algorithm of characteristics of body surface of infected fish,” *2010 World Automation Congress*, Kobe, 2010, pp. 155-160.
- [15] L. Shi, R. Guo and Y. Ma, “A novel artificial fish swarm algorithm for pattern recognition with convex optimization,” *2016 International Conference on Communication and Electronics Systems (ICCES)*, Coimbatore, 2016, pp. 1-4, <https://doi.org/10.1109/CESYS.2016.7889830>
- [16] S. Luo, X. Li, D. Wang, J. Li and C. Sun, “Automatic Fish Recognition and Counting in Video Footage of Fishery Operations,” *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*, Jabalpur, 2015, pp. 296-299, <https://doi.org/10.1109/CICN.2015.66>
- [17] G. Ding et al., “Fish recognition using convolutional neural network,” *OCEANS 2017–Anchorage*, Anchorage, AK, 2017, pp. 1-4.
- [18] A. Krizhevsky; I. Sutskever; G. Hinton. Imagenet classification with deep convolutional neural networks. En *Advances in neural information processing systems*. 2012. pp. 1097-1105.
- [19] M. D. Zeiler, y R. Fergus, “Visualizing and Understanding Convolutional Networks”. *ECCV 2014, Part I, LNCS 8689*, pp. 818–833, 2014. Springer International Publishing Switzerland 2014, https://doi.org/10.1007/978-3-319-10590-1_53
- [20] M.Aursand, I.B. Standal, A. Praël, L. McEvoy, J. Irvine y D.E. Axelson, “¹³C NMR pattern recognition techniques for the classification of atlantic salmon (*salmo salar* l.) according to their wild, farmed, and geographical origin”, *Journal of agricultural and food chemistry*, vol. 57, n° 9, pp. 3444–3451, 2009, <https://doi.org/10.1021/jf8039268>
- [21] M.A. Nanny, R.A. Minear y J.A. Leenheer, “Nuclear magnetic resonance spectroscopy in environmental chemistry”, Oxford University Press, 1997.
- [22] X. Li, M. Shang, H. Qin y L. Chen “Fast accurate fish detection and recognition of underwater images with fast r-cnn”. In *OCEANS’15 MTS/IEEE Washington*, pp. 1–5. IEEE, 2015.
- [23] H. Qin, X. Li, J. Liang, Y. Peng y C. Zhang, “Deepfish: Accurate underwater live fish recognition with a deep architecture” *Neurocomputing*, vol. 187, pp. 49–58, 2016, <https://doi.org/10.1016/j.neucom.2015.10.122>
- [24] L. Jin y H. Liang, “Deep learning for underwater image recognition in small sample size situations” In *OCEANS 2017–Aberdeen*, pp. 1–4. IEEE, 2017, <https://doi.org/10.1109/OCEANSE.2017.8084645>
- [25] P.X. Huang, B.B. Boom y R.B. Fisher “Fish recognition ground-truth data” [En línea], Disponible en: <http://groups.inf.ed.ac.uk/f4k/GROUNDTRUTH/RECOG/>