

Aprendizaje Profundo para la Identificación de Objetos en Robótica Móvil

Olivetto Rendón Alexis Josué¹, Villanueva Tavira Jonathan¹✉, Valdez Martínez Jorge Salvador¹, Magadán Salazar Andrea²

alexisale1944@hotmail.com, ✉jonathanvillanueva@utez.edu.mx, jorgevaldez@utez.edu.mx, magadan@cenidet.edu.mx

¹División Académica de Mecánica Industrial, Universidad Tecnológica Emiliano Zapata
²Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET)

Resumen

En este artículo se muestra la comparación y evaluación de dos redes neuronales artificiales de tipo convolucional (ALEXNET y VGGNET) aplicado a un sistema de reconocimiento automatizado de objetos. Para lograr esto, se programaron las dos topologías de redes neuronales convolucionales en MATLAB, las cuales fueron entrenadas con cuatro categorías de imágenes para obtener como resultados la cuantificación del tiempo de procesamiento, precisión, datos perdidos, tasa de aprendizaje. Esta información sirvió para determinar qué red convolucional presenta mayores beneficios en el proceso de clasificación.

Palabras clave: Aprendizaje profundo, Reconocimiento de objetos, AlexNet, VGGNET, Aprendizaje automático

Abstract

In this paper the comparison and evaluation of convolutional neural networks (AlexNet and VGGNET) were applied to recognize objects. These convolutional neural networks topology were developed in MATLAB and were trained with four different categories of images, obtaining the quantification of execution time, accuracy, data loss, learning rate. This information was used to determine which convolutional network presents the greatest benefits in the classification process

Keywords: Deep learning, Object recognition, AlexNet, VGGNET, Machine learning

1. Introducción

El cerebro es el órgano más increíble del cuerpo humano. Interpreta la forma en que se percibe cada imagen, sonido, olor, sabor y tacto. Permite almacenar recuerdos, experimentar emociones e incluso soñar. Durante décadas, se ha tratado de construir máquinas inteligentes con cerebros, como asistentes robotizados para limpiar casas, coches que se conducen solos, microscopios que detectan enfermedades automáticamente, pero construir estas máquinas artificialmente inteligentes obliga a resolver algunos de los problemas computacionales más complejos con los que se han luchado, problemas que el cerebro humano puede resolver de una determinada manera en microsegundos [1].

En las últimas décadas, el *machine learning* se ha convertido en uno de los pilares de la tecnología de la información. Con la creciente cantidad de datos disponibles hay buenas razones para creer que el análisis inteligente de datos se hará aún más omnipresente como un ingrediente necesario para el progreso tecnológico [2].

Nuevas técnicas como aprendizaje profundo (basado en redes neuronales) han surgido para tratar de simular en una máquina la forma en el que un cerebro funciona y aprende. El aprendizaje profundo es una técnica que se deriva del aprendizaje automático o mejor conocido como *Machine Learning*. El aprendizaje automático utiliza varias capas de procesamiento no lineal para aprender representaciones útiles de características directamente a partir de los datos. Los modelos de aprendizaje automático pueden lograr una precisión muy alta en la clasificación de objetos [3]. Esta técnica consiste en desfragmentar en profundidad diferentes características de un elemento de información y de esta manera aprender por medio de patrones, que no son evidentes de reconocer para una máquina. La motivación para el uso de las técnicas de aprendizaje profundo ha surgido como una posibilidad de tratar las abstracciones de alto nivel de una manera más eficiente, como el caso del reconocimiento de las personas en diferentes posiciones o situaciones, el aprendizaje profundo ha sido la nueva frontera del aprendizaje automático que se ha aplicado a diversos problemas complejos de reconocimiento y modelado de patrones, es decir, la clasificación de objetos, video e imagen, etc., con eficacia [4].

Una de las partes más importantes de una red neuronal artificial es el aprendizaje. El esquema del aprendizaje de una red es lo que determina el tipo de problemas que será capaz de resolver. Las redes neuronales artificiales son sistemas de aprendizaje basados en ejemplos. La capacidad de una red para resolver un problema estará ligada de forma fundamental al tipo de ejemplos [5], por lo que, en este trabajo, se ha propuesto y evaluado un sistema de reconocimiento automatizado de objetos basado en redes convolucionadas utilizando un conjunto de redes neuronales artificiales para comparar el desempeño de cada una de ellas en el software comercial MATLAB.

2. Motivación

En la División Académica de Mecánica Industrial (DAMI) de la Universidad Tecnológica Emiliano Zapata del Estado de Morelos se han desarrollado diversos proyectos relacionados con la robótica móvil. Sin embargo, hasta ahora los robots desarrollados han sido programados con algoritmos de *Hard Computing* en lenguajes como: Robot C, LabVIEW, Arduino, Python y MATLAB. Es importante destacar que cada uno de los robots diseñados han seguido la metodología del Dr. Richard Grimmett que establece que para diseñar correctamente robots móviles estos deben cumplir con tres fases: a) Contar con dispositivos que permitan al robot la posibilidad de desplazarse correctamente, b) los robots deben de ser dotados de sentidos que le permitan ver a través de una cámara, escuchar por medio de un micrófono y hablar mediante una bocina; finalmente, el tercer paso es dotar a estos de una etapa de inteligencia que les permita evadir obstáculos, reconocer objetos y optimizar sus trayectorias [6]. Con este tipo de trabajos en un corto plazo se busca dotar de identificación de objetos empleando la técnica de aprendizaje profundo.

3. Objetivo general

Realizar un estudio comparativo con dos tipos de redes neuronales convolucionales aplicadas al aprendizaje profundo que permitan la clasificación de objetos en un sistema de robótica móvil.

3.1 Objetivos particulares

- Analizar dos tipos de redes neuronales convolucionales aplicadas al aprendizaje profundo en el software MATLAB.
- Comparar los resultados obtenidos para cada uno de los modelos de redes neuronales artificiales convolucionales.
- Cargar la red neuronal artificial entrada en un Sistema embebido como Arduino, Raspberry Pi o la tarjeta myRIO.

4. Redes neuronales convolucionales

Una red neuronal artificial convolucional es una técnica muy común en aprendizaje profundo, inspirada en el mecanismo de percepción visual de las criaturas vivas [7]. En esencia son redes neuronales artificiales que emplean la operación de la convolución como una de sus capas. Actualmente, este tipo de redes se han aplicado exitosamente en problemas donde los datos de entrada sobre los cuales se van a realizar las predicciones cuentan con una reja conocida como topología de una serie temporal o una imagen [8]. A continuación, se describe el proceso de convolución de este tipo de redes neuronales artificiales [9].

Primero se desarrolla la operación de convolución en una dimensión. Dada una entrada $I(t)$ y un kernel $K(a)$ la operación de convolución es dada por (1).

$$s(t) = \sum_a I(a)K(t-a) \quad (1)$$

Sin embargo, una forma equivalente de esta operación dada la conmutatividad de la operación de la convolución está dada por la siguiente expresión (2).

$$s(t) = \sum_a I(t-a)K(a) \quad (2)$$

Finalmente, el signo negativo puede ser sustituido para obtener la correlación cruzada como sigue en (3).

$$s(t) = \sum_a I(t+a)K(a) \quad (3)$$

En la literatura sobre aprendizaje profundo y las implementaciones de software, la convolución y la correlación cruzada se utilizan indistintamente. La esencia de la operación es que el kernel es un conjunto de datos mucho más pequeño en comparación con la entrada, y la salida de la operación de convolución es mayor cuando la entrada es similar al núcleo. Las figuras 1 y 2 ejemplifican mejor esta idea en la que se toma una entrada aleatoria y un kernel arbitrario, se realiza la operación de convolución y el valor más alto se logra cuando el kernel es similar a una porción particular de entrada [9].

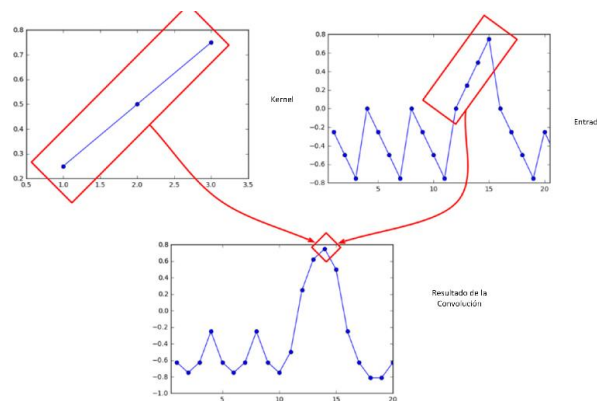


Figura 1. Operación de convolución intuitiva [9].

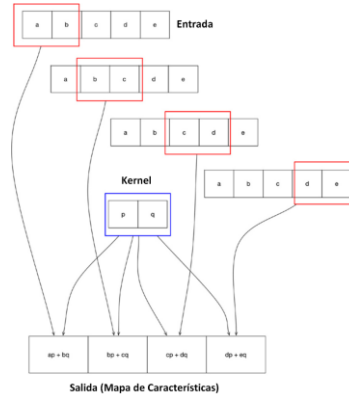


Figura 2. Operación de convolución de una dimensión [9].

De las anteriores figuras se puede observar los siguientes puntos:

- La entrada es un gran conjunto de datos arbitrarios.
- El kernel es el conjunto de datos más pequeño que el número de datos de entrada.
- La operación de convolución se realiza con los datos del kernel y la entrada calculando la similitud de los mismos.
- La operación de convolución produce el valor más alto donde el kernel se asemeja más a una proporción de los datos de entrada.

La operación de convolución puede ser extendida a dos dimensiones (Ver figura 3), dada una entrada $I(m,n)$ y un kernel $K(a,b)$, como se ve en (4).

$$s(t) = \sum_a \sum_b I(a,b)K(m-a,n-b) \quad (4)$$

Una equivalencia de esta operación dada la conmutatividad de la operación de la convolución está dada por la siguiente expresión en (5).

$$s(t) = \sum_a \sum_b I(m-a,n-b)K(m,n) \quad (5)$$

Finalmente, se puede reemplazar el signo negativo para obtener una correlación cruzada lo que arroja la expresión (6).

$$s(t) = \sum_a \sum_b I(m+a,n+b)K(m,n) \quad (6)$$

Finalmente, después de mencionar la operación de convolución, ahora se puede profundizar más en las partes que constituyen a las redes neuronales convolucionales, cuya característica principal es el uso de una capa de convolución en lugar de una capa totalmente conectada que implique una multiplicación de matriz. Por lo tanto, una capa totalmente conectada puede ser descrita de la siguiente forma $y = f(x, w)$ donde x es el vector de entrada, y el de salida, w es un conjunto de pesos y finalmente, f es la función de activación. Adicionalmente, una capa de convolución se puede describir cómo $y = f(s(x, w))$ donde s denota la operación de convolución entre la entrada y los pesos [9].

En la figura 4 se muestra una red neuronal con una capa completamente conectada, mientras que la figura 5 muestra una red que realiza el proceso de la convolución.

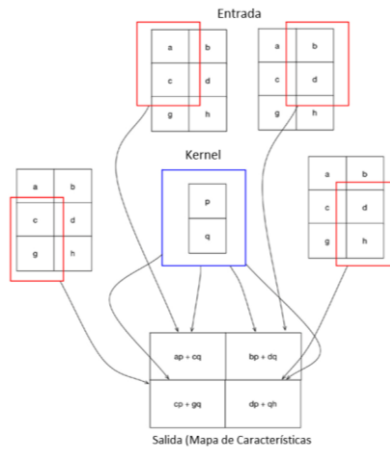


Figura 3. Operación de convolución de dos dimensiones [9].

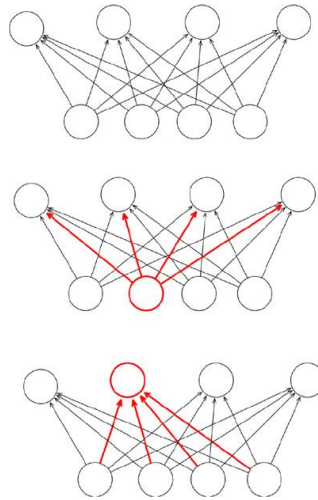


Figura 4. Operación de convolución de dos dimensiones [9].

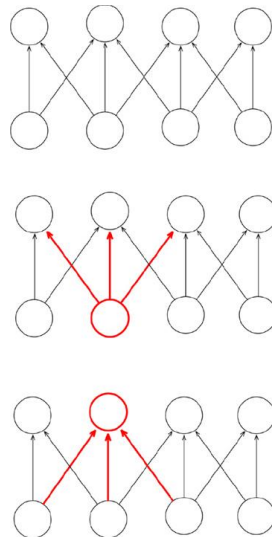


Figura 5. Interacciones dispersas en la capa de convolución [9].

Finalmente, en la figura 6 se observa el comportamiento de los parámetros en una capa de convolución y la falta de ella en una capa completamente conectada. De lo anterior, se puede señalar lo siguiente.

- Para el mismo número de entradas y salidas, la capa conectada contiene más conexiones y correspondientemente una capa contiene los pesos para la convolución.
- Los resultados que se producen empleando la capa de convolución conllevan un menor número de iteraciones con respecto a una capa que se encuentra totalmente conectada.
- Los parámetros y los pesos se comparten a través de la capa de convolución, dado que el núcleo es mucho más pequeño que la entrada y el kernel se desliza a través de la entrada.

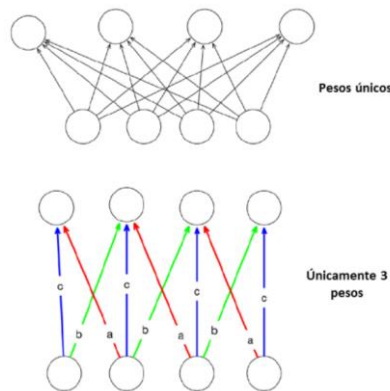


Figura 6. Parámetros compartidos de pesos. [9]

A continuación, se describen los dos tipos de redes neuronales convolucionales empleadas en este trabajo de investigación.

4.1 ALEXNET

Esta red neuronal de tipo convolucional fue creada inicialmente para un concurso llamado *ImageNet Large Scale Visual Recognition Competition (ILSVRC)*, que es una competencia considerada como los juegos olímpicos de la visión por computadora, donde diversos equipos de varias partes del mundo participan para encontrar el mejor modelo de visión por computadora para tareas como clasificación, localización y detección. *Alexnet* está formada por cinco capas convolucionales, capas de agrupación máxima, capas eliminadas, y tres capas totalmente conectadas. La red se diseñó para la clasificación de mil posibles categorías [10].

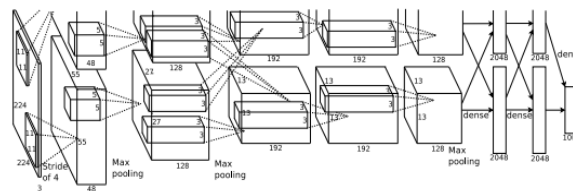


Figura 7. Arquitectura de la Red AlexNet [9].

4.2 VGGNET

Es una red de 16 capas utilizada por el grupo de Geometría Visual de la Universidad de Oxford con el que compitieron en el concurso llamado *ImageNet Large Scale Visual Recognition Competition*

(ILSVRC). Esta red contiene una base de entrenamiento de 254*254 imágenes en RGB a través de 5 bloques de capas convolucionales donde cada bloque está compuesto por un número creciente de filtros de 3x3.

5. Procedimiento

Para realizar las pruebas de las dos diferentes topologías de redes neuronales convolucionales se empleó el software comercial MATLAB en su versión 2017. Las redes fueron entrenadas para clasificar cuatro categorías de imágenes como: sillas, ventiladores, switches y silla escolar. Es importante mencionar que cada red como entrada debe de contener imágenes en diferentes tamaños, para el caso de *AlexNet* se requieren imágenes de 227x277 pixeles y en el caso de *VGG16* se requieren 224x224. En la figura 8 se observan un conjunto de imágenes empleadas para el entrenamiento de estas dos topologías.

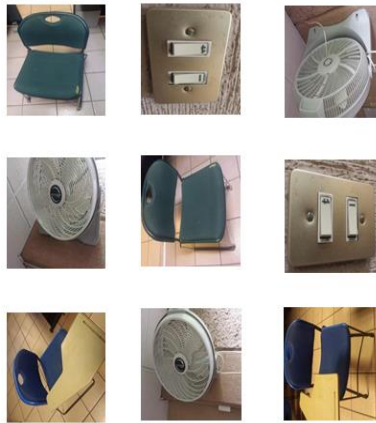


Figura 8. Conjunto de imágenes para entrenamiento.

A continuación, en la figura 9, se muestra el fragmento de código empleado en MATLAB para el entrenamiento de las redes neuronales convolucionales, así como en la figura 10 las pantallas de ejecución para la etapa de reconocimiento.

```
clear cámara;  
camara=webcam;  
redcov=Mynet;  
  
while true  
    capt=camera.snapshot;  
    capt=imresize(capt, [227,227]);  
    label=classify(redcov,picture);  
    image(capt);  
    title(char(label));  
    drawnow;  
end
```

Figura 9. Código en MATLAB para la etapa de entrenamiento.

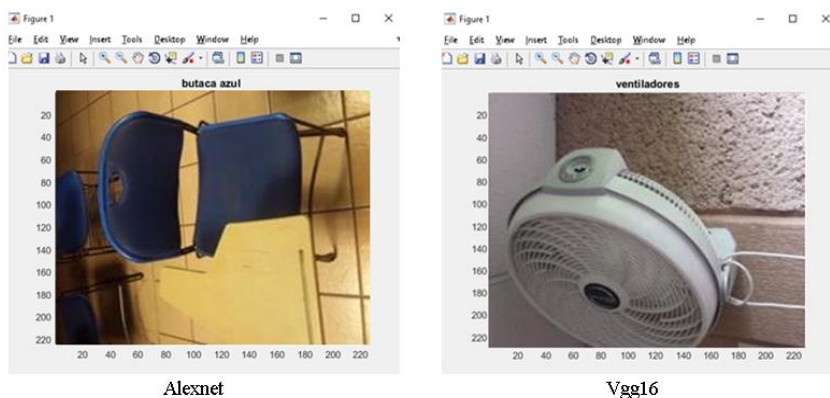


Figura 10. Vista del funcionamiento de la cámara y la red neuronal empleada.

6. Resultados

Después de realizar las pruebas correspondientes para cada uno de los modelos de red neuronal convolucional, se obtuvieron los siguientes resultados en la que se pueden observar el tiempo de procesamiento, precisión, datos perdidos, tasa de aprendizaje entre otros, con el objeto de verificar en la práctica que tipo de red neuronal convolucional presenta mayores beneficios en el proceso de clasificación, ver tabla 1 y 2.

Tabla 1. Resultados obtenidos

Redes	Características			
	Épocas	Tiempo de entrenamiento	Pérdida de Datos	Precisión
Alexnet	8	2780 s.	1.275	100%
Vgg16	1	8120 s.	1.7992	22.22%

Tabla 2. Resultados obtenidos.

Redes	Características			
	Capas	Tasa de Aprendizaje	Tipo de imágenes	Tamaño de Imágenes
Alexnet	25	0.0010	2D	227x227
Vgg16	41	0.0010	2D y 3D	224x224

7. Conclusiones

En este trabajo se constató con los resultados de la identificación sobre los objetos deseados, la identificación de los objetos fue de manera distinta para cada red neuronal convolucional, para *AlexNet* el procesamiento y la identificación del objeto fue más rápido que el de *VGG16*, esto es debido a la gran cantidad de capas que contienen cada una y esto afecta el tiempo para la obtención de una respuesta. En estas pruebas no se ocuparon unidades de procesamiento grafico (*GPU*), estos pueden agilizar el proceso ya que el tiempo de procesamiento y entrenamiento de un modelo de aprendizaje profundo puede llevar mucho tiempo, desde días hasta semanas. El uso de la aceleración mediante *GPU* puede agilizar este proceso y reducir el tiempo.

Con los resultados adquiridos, se llegó a la conclusión que para esta problemática es más viable ocupar la red neuronal convolucional *AlexNet*, debido a su procesamiento más eficaz y su entrenamiento más rápido en comparación a la red neuronal convolucional *VGG16*.

Referencias

- [1] Buduma N. y Locascio N. *Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms*, "O'Reilly Media, Inc. Pp. 228. 2006.
- [2] Smola A. y Vishwanathan S.V.N.. *Introduction to machine learning*. Ed. Cambridge University Press. 2008
- [3] De la Rosa E. El aprendizaje profundo para la identificación de sistemas no lineales, Tesis de Maestría Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional. 2014.
- [4] De Marsico M., Petrosino A. y Ricciardi S. Iris Recognition through Machine Learning techniques: a survey, *Pattern Recognition Letters 82(2)*. Pp. 106 – 115. 2016.
- [5] Isasi P. y Galván I. *Redes de Neuronas Artificiales un Enfoque Práctico*, Ed. Pearson. Pp. 229. 2014.
- [6] Grimmert R. "The New Accessible Robotics: Building Amazing Projects with Inexpensive Hardware and Open Source Software", International Conference on Mechatronics, Electronics and Automotive Engineering, 2014.
- [7] Tindall L., Luong C. y Saad A. Plankton Classification Using VGG16 Network. Noiselab University of California. 2017
- [8] Gu J., Wang Z., Kuen J., Ma L. Recent Advances in convolutional neural networks, *Pattern Recognition 77*. Pp. 354 – 377. 2017
- [9] Ketkar N. *Deep Learning with Python: A Hands-on Introduction*, Apress, Pp. 226. 2017.
- [10] Krizhevsky A., Sutskever I. y Hinton G. E. (). "Image Net Classification with Deep Nonvolutional Neural Networks", NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems – Vol. 1. 2012