



Reconocimiento de objetos del hogar, usando redes neuronales convolucionales para personas con discapacidad visual

Recognition of household objects, using convolutional neural networks for people with visual impairment

Reconhecimento de objetos domésticos, usando redes neurais convolucionais para pessoas com deficiência visual

Jaime Israel Izquierdo-Valladarez ^I

ijaime@psg.ucacue.edu.ec

<https://orcid.org/0000-0002-4726-4201>

Juan Pablo Cuenca-Tapia ^{II}

jcuenca@ucacue.edu.ec

Correspondencia: ijaime@psg.ucacue.edu.ec

Ciencias de las ingenierías

Artículo de investigación

***Recibido:** 22 de noviembre de 2019 ***Aceptado:** 30 diciembre de 2019 * **Publicado:** 17 de enero 2020

- I. Ingeniero en Electrónica y Telecomunicaciones, Dependencia de Posgrados, Docente de la Universidad Católica de Cuenca, Cuenca, Ecuador.
- II. Ingeniero Electrónico, Docente de la Unidad Académica de Ingeniería, Industria y Construcción de la Universidad Católica de Cuenca, Jefatura de Posgrados, Cuenca, Ecuador.

Resumen

Las personas con dificultades visuales tienen múltiples necesidades en su vida diaria. En muchas ocasiones, ellos necesitan de terceras personas para lograr ciertas actividades que requieren al momento de movilizarse. Hoy en día, múltiples soluciones tecnológicas han sido desarrolladas para ayudar a personas con deficiencia visual. El presente artículo, muestra las etapas de entrenamiento de una red neuronal convolucional (CNN) para el reconocimiento de imágenes de objetos del hogar, usando la red AlexNet y MATLAB. Para la etapa del entrenamiento se usarán 14 tipos de objetos del hogar: lavadora, televisor, sofá, sombrilla, grapadora, balón soccer, tijera, laptop, lámpara, jarro, taza, silla, cellphone y cámara. Se plantean los siguientes pasos para el entrenamiento de la red neuronal convolucional: recolección de las imágenes para entrenamiento, procesamiento de las imágenes, entrenamiento de la red neuronal en base a las imágenes y validación satisfactoria del modelo. Las imágenes son obtenidas de páginas digitales, se las organiza por su clase o nombre, para que sean utilizadas en el modelo de entrenamiento. La validación se la realiza calculando la fracción de imágenes de prueba correctamente clasificadas, con lo cual se determina la eficacia del sistema entrenado. Los resultados de la presente investigación, tienen un alcance exponencial en los sistemas autónomos dedicados con visión artificial, usando bajo recurso computacional, cuyo objetivo es demostrar que el reconocimiento de imágenes dentro del hogar aporta una ayuda valiosa a las personas con discapacidad visual.

Palabras clave: Aprendizaje de máquina; visión artificial; redes neuronales convolucionales; Alexnet; MatLab.

Abstract

People with visual difficulties have multiple needs in their daily lives. In many occasions they need to use another person to attaches some activities that they require at the moment walking around. Today multiple solutions have been developed to assist people with visual impairment. This paper shows the steps of convolutional neural network learning image recognition of home objects, using AlexNet network and MATLAB. To the training stage, 14 types of household objects have been used, washing machine, TV, sofa, umbrella, stapler, soccer ball, scissors, laptop, lamp, jug, cup, chair, cellphone, and camera. The following steps are proposed for the training of the convolutional neural network: collection the images for training, image

processing, training of neuronal network based on the images and satisfactory validation of the model. Images are obtained through digital pages; it is organized by label names or class in order to be used in training model. Validation is done by calculating the fraction of correctly classified test images, which determines the effectiveness of the trained system. The results of this investigation have an exponential reach dedicated autonomous systems with artificial vision using low computational resources and whose objective is to demonstrate that recognition of images inside home contribute a valuable help to people with visual impairment.

Keywords: Machine learning; artificial vision; convolutional neural network; AlexNet; MatLab.

Resumo

Pessoas com dificuldades visuais têm múltiplas necessidades em suas vidas diárias. Em muitas ocasiões, eles precisam de terceiros para realizar determinadas atividades necessárias no momento da mobilização. Hoje, várias soluções tecnológicas foram desenvolvidas para ajudar pessoas com deficiência visual. Este artigo mostra as etapas de treinamento de uma rede neural convolucional (CNN) para o reconhecimento de imagens de objetos domésticos, usando as redes AlexNet e MATLAB. Para a etapa de treinamento, serão utilizados 14 tipos de objetos domésticos: máquina de lavar roupa, TV, sofá, guarda-chuva, grampeador, bola de futebol, tesoura, laptop, abajur, jarro, copo, cadeira, celular e câmera. As etapas a seguir são propostas para o treinamento da rede neural convolucional: coleta das imagens para treinamento, processamento de imagens, treinamento da rede neural com base nas imagens e validação satisfatória do modelo, obtidas em páginas digitais, eles são organizados por classe ou nome, para serem usados no modelo de treinamento. A validação é feita calculando a fração de imagens de teste classificadas corretamente, o que determina a eficácia do sistema treinado. Os resultados da presente investigação têm um alcance exponencial nos sistemas autônomos dedicados à visão artificial, utilizando recursos computacionais, cujo objetivo é demonstrar que o reconhecimento de imagens dentro de casa fornece uma ajuda valiosa para pessoas com deficiência visual.

Palavras-chave: Aprendizado de máquina; visão artificial; redes neurais convolucionais; Alexnet; MatLab.

Introducción

La Organización Mundial de la Salud (OMS) estima que más de mil millones de la población mundial sufren de algún tipo de discapacidad. Según informe emanado por este organismo: “del total estimado de personas con discapacidad 110 millones (2,2%) tienen dificultades muy significativas de funcionamiento” OMS (2011), mientras que “la Carga Mundial de Morbilidad cifra en 190 millones (3,8%) las personas con una discapacidad grave” OMS, (2018). Entre éstas, la discapacidad visual está presente en aproximadamente 1300 millones de personas que afecta principalmente en su vida social.

Ante este panorama, el avance de la ciencia y la tecnología en la actualidad ha permitido desarrollar aplicaciones de asistencia social, para personas con discapacidad visual, mediante el uso computacional de la visión artificial. A este respecto, la visión artificial ha permitido la evolución del reconocimiento de imágenes basado en redes neuronales. Según (Haykin, 1994):

Una red neuronal es una máquina capaz de modelar la forma en que el cerebro realiza una tarea o función particular de interés. La red se la puede implementar con componentes electrónicos o es simulada en software sobre un computador digital.

De acuerdo con esta afirmación, al reconocer eficazmente una imagen y procesarla conlleva un gran aporte para la movilidad independiente de una persona sin buena visión.

Partiendo de lo afirmado el objetivo de esta investigación es entrenar un modelo de visión artificial utilizando una red neuronal convolucional para reconocimiento de objetos dentro del hogar, los cuales serán: lavadora, televisor, sofá, sombrilla, grapadora, balón soccer, tijera, laptop, lámpara, jarro, taza, silla, cellphone y cámara.

Desarrollo

Los ojos son el principal órgano de la visión. De acuerdo con Zhu (2001) “La perfección del ojo es tal, que cada ojo ha evolucionado adaptándose a las necesidades del organismo poseedor, lo que ha provocado que existan diversas diferencias en la anatomía y fisiología ocular de los diferentes organismos.” Así, La estructura más simple de ojo consiste en una concentración de células foto receptoras mediante las cuales un organismo puede distinguir, no sólo la luz y la oscuridad, sino también la dirección de la luz incidente.

Sin embargo, el sistema óptico complejo presente en el 96 % de las especies animales, es capaz de realizar un proceso completo que comienza con la detección de la luz y finaliza con unos impulsos electroquímicos viajando a través de las neuronas. Durante ese proceso, los ojos tienen que captar la luz, regular la intensidad mediante un diafragma y, mediante un sistema de lentes (cristalino), enfocarla en único punto que se encargará de realizar la transformación en impulsos eléctricos. Este punto donde convergen todos los rayos de luz, que será objeto de estudio durante este trabajo, es conocido como retina.

A nivel microscópico, la retina tiene una estructura compleja formada por varias capas de neuronas interconectadas. Existen dos tipos principales de foto receptores en la retina: los conos y los bastones. “Las células de la retina presentan grandes similitudes con las del cerebro, apoyando la afirmación común de que el sistema visual es una extensión del sistema nervioso central.” Zhu (2001).

Para el reconocimiento de imágenes mediante el uso de la tecnología para personas con discapacidad visual, se hace uso de las redes neuronales convolucionales (CNNs por sus iniciales en inglés), “las cuales son un modelo donde las neuronas corresponden a campos receptores de una manera muy similar a las neuronas de la corteza visual primaria de un cerebro biológico.” Goodfellow (2012). Según Cireşan, (2010). “el éxito de estos sistemas se debe en gran parte a la evolución de la capacidad computacional.” Por otro lado, las redes neuronales artificiales (ANN, Artificial Neural Networks) y actualmente redes neuronales profundas (DNN, Deep Neural Networks)

Son modelos matemáticos que simulan el comportamiento del cerebro humano y han sido ampliamente utilizadas en tareas de clasificación como, por ejemplo: en la detección de imágenes de percepción remota, la identificación de fraudes bancarios o el diagnóstico de enfermedades, entre otras. Monroy y otros (2017)

En este sentido, el nombre de red neuronal convolucional indica que la red emplea una operación matemática llamada convolución. Según Goodfellow, Bengio y Courville (2016) “Convolución es un tipo de especialización de operación lineal, la cual es usada en al menos una de las capas de las redes neuronales.”

Ecuación 1

$$s(t) = \int x(a)w(t - a) da$$

Fuente: Goodfellow (2016)

La operación de la ecuación 1 es llamada convolución, esta operación es típicamente denotada como asterisco como se observa en la ecuación 2.

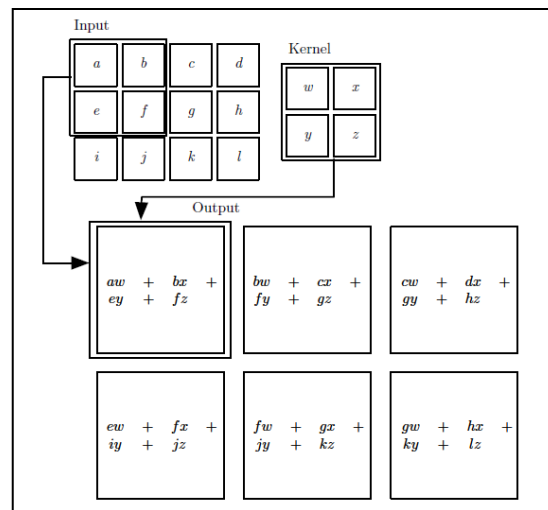
Ecuación 2

$$s(t) = (x * w)(t)$$

Fuente: (Goodfellow et al., 2016)

En la ecuación 2, se pueden observar dos argumentos, el primero hace referencia a input y el segundo es referido como kernel. La salida es muchas veces denotada como feature map.

Figura 2. Ejemplo de aplicación de convolucional

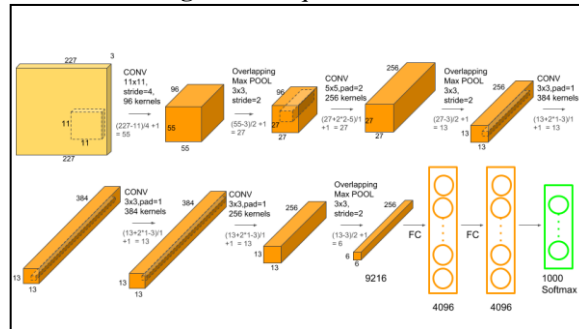


Fuente: (Goodfellow et al., 2016)

En este mismo orden de ideas, existe una gran cantidad de arquitecturas de redes convolucionales, que han demostrado ser eficaces en diversos campos. Una de ellas, “la arquitectura de AlexNet consiste de cinco capas de convolución y tres capas de conexión completa.” (Nayak, 2018). Otra característica es que la imagen de entrada tiene que tener un tamaño de 227x227. Las dos primeras capas de convolución son seguidas de capas overlapping max pooling, las cuales son usadas para reducir el ancho y alto, pero mantienen la profundidad. Luego de cada convolución existe la función ReLU, la cual se encarga de realizar una rectificación lineal. De la tercera capa a la quinta están conectadas directamente. Luego de la quinta capa, están conectada a un overlapping max layer, las cuales van a dos series de fully

connected layers. Finalmente, estas dos están conectadas a softmax classifier con 1000 clases de etiquetas. Loncomilla, (2016)

Figura 3. Arquitectura de AlexNet



Fuente: <https://medium.com/@eltronicsvilla17/convolutional-neural-network-1a02f472a90c>

Metodología

En esta investigación se realiza el siguiente procedimiento:

Recolección de las imágenes de categorías de objetos a entrenar, se obtienen imágenes mediante dispositivos digitales y descargas desde el internet con buena calidad de resolución, las cuales serán organizadas dentro de una carpeta y organizándolas según su clase o categoría.

Procesamiento de imágenes, se utilizará la herramienta MATLAB para procesar las imágenes de interés previo al entrenamiento.

Entrenamiento de la red neuronal convolucional, se utiliza el software MATLAB y la red AlexNet, realizaremos transferencia de aprendizaje para utilizar nuestras imágenes de procesamiento como nuevo aprendizaje.

Validación del modelo entrenado, se usa una base de datos de imágenes para validar el funcionamiento del modelo entrenado, se analiza el porcentaje de eficiencia mediante la comparación de imágenes correctamente clasificadas

Resultados

El entrenamiento del aprendizaje de la CNN se ejecuta mediante el software MatLab usando el modelo modificado de la red AlexNet.

Recolección de imágenes objetos del hogar

Figura 4. Clases de objetos del hogar



Elaboración: Autoría propia

Las 14 clases de imágenes que se muestran en la figura 4 se las obtuvo del recurso libre de internet (google) y se las organizó en una carpeta según su nombre. El formato de las imágenes es JPG para ser utilizadas en el modelo de entrenamiento de la CNN. Se obtuvo un total de 1120 imágenes de las 14 clases de objetos a entrenar, como se muestra en la tabla 1.

Tabla 1. Cantidad de las clases de imágenes

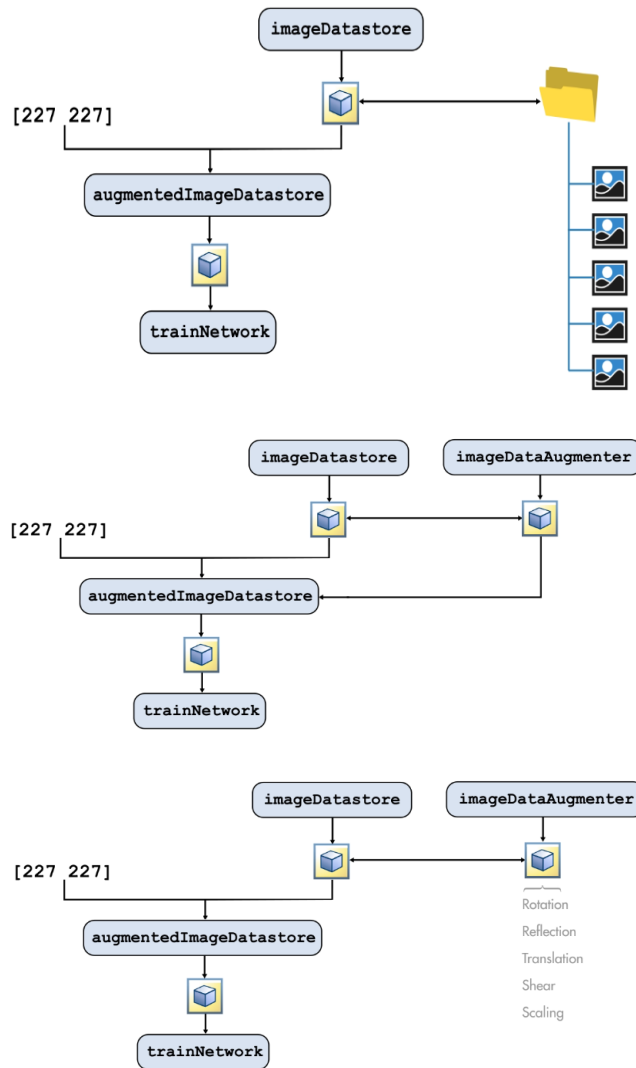
Clase	Cantidad	Porcentaje
Lavadora	80	7,14%
Televisor	80	7,14%
Sofá	80	7,14%
Sombrilla	80	7,14%
Grapadora	80	7,14%
Balón Soccer	80	7,14%
Tijera	80	7,14%
Laptop	80	7,14%
Lámpara	80	7,14%
Silla	80	7,14%
Cellphone	80	7,14%
Jarro	80	7,14%
Taza	80	7,14%
Cámara	80	7,14%
Total	1120	100%

Fuente: Autoría propia

Procesamiento de imágenes

Las imágenes se las arregla categóricamente en carpetas ya que las etiquetas se leen automáticamente desde los nombres de la carpeta. Luego con la herramienta MatLab se realiza una imagen (augmentedImageDatastore) de la data, previo a ser entrenada, como se muestra en la figura 5.

Figura 5. Proceso previo al entrenamiento de la CNN



Fuente: Matlab

Entrenamiento de la red neuronal convolucional

Para el entrenamiento utilizaremos AlexNet y la programación en MatLab para el aprendizaje de las 14 categorías de objetos de hogar. El aprendizaje es ejecutado en un computador i7-6500CPU @ 2.5GHZ, 12GB RAM, 1000GB HD.

Antes del entrenamiento, realizamos la transferencia de aprendizaje para lo cual modificamos la capa de clasificación (capa 23ava) y salida (capa 25ava) de la arquitectura de AlexNet, con lo cual prácticamente realizamos una modificación de la red como se observa en la figura 6.

Figura 6. Capas a modificar de AlexNet

Modify the classification and output layers

```
layers(end-2) = fullyConnectedLayer(numClasses);
layers(end) = classificationLayer;
```

Fuente: Autoría propia

La figura 7 se observa las capas que contiene la arquitectura AlexNet.

Figura 7. Resultado de las capas de AlexNet

Layer	Type	Description
1	'data'	Image Input
2	'conv1'	Convolution
3	'relu1'	ReLU
4	'norm1'	Cross Channel Normalization
5	'pool1'	Max Pooling
6	'conv2'	Grouped Convolution
7	'relu2'	ReLU
8	'norm2'	Cross Channel Normalization
9	'pool2'	Max Pooling
10	'conv3'	Convolution
11	'relu3'	ReLU
12	'conv4'	Grouped Convolution
13	'relu4'	ReLU
14	'conv5'	Grouped Convolution
15	'relu5'	ReLU
16	'pool5'	Max Pooling
17	'fc6'	Fully Connected
18	'relu6'	ReLU
19	'drop6'	Dropout
20	'fc7'	Fully Connected
21	'relu7'	ReLU
22	'drop7'	Dropout
23	'fc8'	Fully Connected
24	'prob'	Softmax
25	'output'	Classification Output

En la figura 8 se muestra la red modificada luego de aplicar el código de la figura 6, lo cual nos indica que está lista para aplicar el entrenamiento.

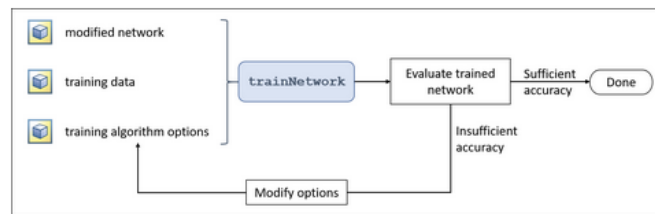
Figura 8. Modificación de Red para utilizarla en CNN

```

25 'output' Classification Output crossentropyex with '1'
layers =
25x1 Layer array with layers:
 1 'data' Image Input 227x227x3 images with 'zei
 2 'conv1' Convolution 96 11x11x3 convolutions w
 3 'relu1' ReLU ReLU
 4 'norm1' Cross Channel Normalization cross channel normalizati
 5 'pool1' Max Pooling 3x3 max pooling with stri
 6 'conv2' Grouped Convolution 2 groups of 128 5x5x48 coi
 7 'relu2' ReLU ReLU
 8 'norm2' Cross Channel Normalization cross channel normalizati
 9 'pool2' Max Pooling 3x3 max pooling with stri
10 'conv3' Convolution 384 3x3x256 convolutions v
11 'relu3' ReLU ReLU
12 'conv4' Grouped Convolution 2 groups of 192 3x3x192 co
13 'relu4' ReLU ReLU
14 'conv5' Grouped Convolution 2 groups of 128 3x3x192 co
15 'relu5' ReLU ReLU
16 'pool5' Max Pooling 3x3 max pooling with stri
17 'fc6' Fully Connected 4096 fully connected layer
18 'relu6' ReLU ReLU
19 'drop6' Dropout 50% dropout
20 'fc7' Fully Connected 4096 fully connected layer
21 'relu7' ReLU ReLU
22 'drop7' Dropout 50% dropout
23 '' Fully Connected 12 fully connected layer
24 'prob' Softmax softmax
25 '' Classification Output crossentropyex
    
```

El proceso de entrenamiento de la CNN básicamente necesita tres componentes: el primero es la red modificada, que en nuestro caso va a ser la red preexistente AlexNet. El segundo componente, son los almacenes de datos previamente etiquetados categóricamente, para utilizarlos como datos de entrenamiento, y finalmente se necesita una variable que contendrá las opciones de entrenamiento. En la figura 9 se muestra el modelo de entrenamiento planteado.

Figura 9. Modelo de entrenamiento de CNN



Fuente: Matlab

Para la etapa del entrenamiento usamos el 70% del augmentedImageDatastore y el 30% restante serán usados para la etapa de prueba de la CNN, la selección de las imágenes las realiza de forma aleatoria.

Evaluación del modelo de la red neuronal convolucional

Tras el entrenamiento de la CNN, con la herramienta MatLab se observa el registro del rendimiento de la red con los datos de entrenamiento en cada iteración, la cual está encapsulada en la variable info, ver figura 10.

Figura 10. Variable info

Perform training

```
[objectshomenet,info] = trainNetwork(trainImgs, layers, options);
```

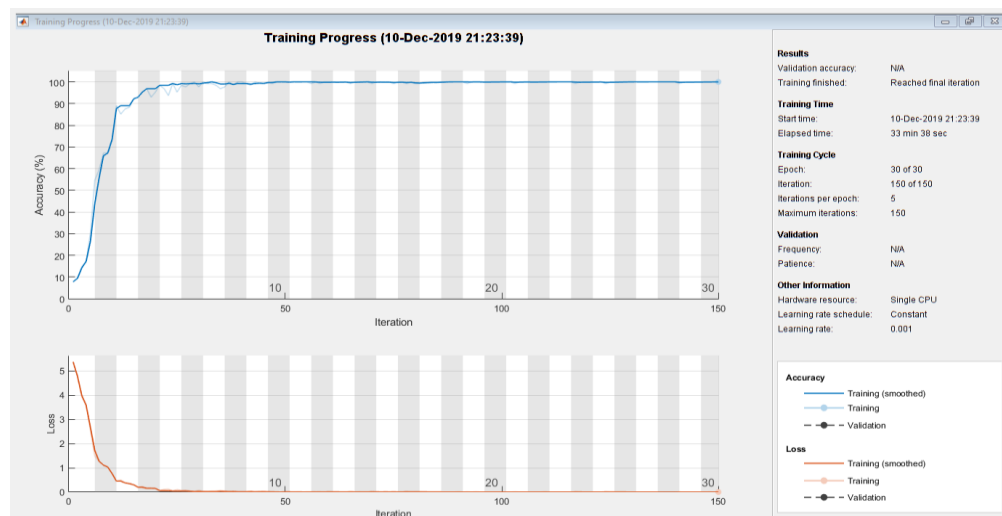
Fuente: Autoría propia

Como se observa en la figura 11 se obtiene un registro de TrainingLoss, TrainingAccuracy y BaseLearnRate.

Figura 11. Entrenamiento de la red

Training on single CPU.
Initializing input data normalization.

Epoch	Iteration	Time Elapsed (hh:mm:ss)	Mini-batch Accuracy	Mini-batch Loss	Base Learning Rate
1	1	00:00:14	7.81%	5.3747	0.0010
10	50	00:09:06	100.00%	0.0142	0.0010
20	100	00:19:44	100.00%	0.0053	0.0010
30	150	00:33:38	100.00%	0.0018	0.0010



Fuente: Autoría propia

Para la evaluación de la CNN luego de la etapa de aprendizaje utilizamos un test con el 30% restante de imágenes, ver figura 12.

Figura 12. Código para test de la CNN

Use the trained network to classify test images

```
testpreds = classify(objectshomenet,testImgs);
```

Fuente: Autoría propia

En la figura 13, se muestra categóricamente la predicción realizada.

Figura 13. Predicción del test

```
testpreds = 448x1 categorical array  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer  
BalonSoccer
```

Fuente: Autoría propia

Evaluación del rendimiento de la red neuronal convolucional

Para determinar cuántas de las imágenes de prueba han sido clasificadas correctamente, se realiza una comparación de la clasificación predicha con la clasificación conocida.

Figura 14. Código para evaluar el rendimiento de la CNN

Evaluate the results

Calculate the accuracy

```
imageActual = testImgs.Labels;  
numCorrect = nnz(testpreds == imageActual)
```

```
numCorrect = 436
```

```
fracCorrect = numCorrect/numel(testpreds)
```

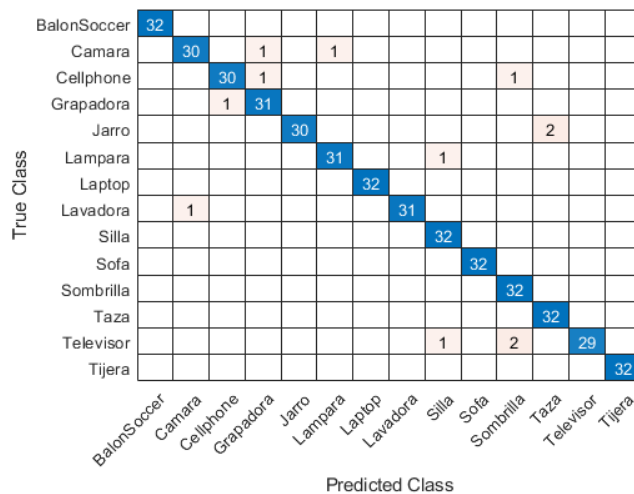
```
fracCorrect = 0.9732
```

Fuente: Autoría propia

Mediante el código de la figura 14 podemos obtener el número correcto de precisión al evaluar los resultados, en el presente caso de obtuvo 436 aceraciones con una fracción correcta de 97.32% de precisión.

A continuación en la figura 15 por medio de confusionchart(testImgs.Labels,testpreds), podemos obtener la matriz de confusión por medio de la cual podemos observar las predicciones erróneas (elementos no diagonales) y también las predicciones correctas (elementos diagonales).

Figura 15. True Class vs Predicted Class



Fuente: Autoría propia

Conclusiones

Para obtener un buen promedio de eficiencia del modelo neuronal, debemos utilizar imágenes con buena resolución. Se logró obtener un 97.32 % de eficiencia en esta investigación.

Se debe reducir el tamaño de las imágenes a 227x227 para obtener un buen rendimiento de aprendizaje.

La cantidad o porcentaje para el aprendizaje y el test debe seleccionarse de manera que su efectividad sea la más óptima, en nuestro caso utilizamos una relación 70/30, para aprendizaje y test respectivamente.

El tiempo de entrenamiento fue de 33 min con 38 segundos realizado en hardware CPU, sin embargo, si se dispone de GPU el tiempo de entrenamiento se reduciría considerablemente.

Con la evaluación obtuvimos 436 aserciones de 448 imágenes que fueron clasificadas para el test, con lo cual se logró un 97.32% de precisión.

Las imágenes que se predijeron erróneamente fueron por similitud a otras imágenes, la CNN es capaz de seguir con entrenamiento para mejorar los errores.

La CNN que fue entrenada con imágenes del hogar, nos genera un valor positivo para ser empleado en proyectos de implementación con smartphone, como ayuda para personas con discapacidad visual.

Referencias

1. Deepthi Jain, B., Thakur, S. M., & Suresh, K. V. (2018). Visual Assistance for Blind Using Image Processing. Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing, ICCSP 2018, 499–503. <https://doi.org/10.1109/ICCSP.2018.8524251>
2. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning (Adaptive Computation and Machine learning). Adaptive Computation and Machine Learning Series. <https://doi.org/10.1016/B978-0-12-391420-0.09987-X>
4. Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation* (1st ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.
5. Kacorri, H., Kitani, K. M., Bigham, J. P., & Asakawa, C. (2017). People with visual impairment training personal object recognizers: Feasibility and challenges. *Conference on Human Factors in Computing Systems - Proceedings*, 2017-May, 5839–5849. <https://doi.org/10.1145/3025453.3025899>
6. Krizhevsky, A., Sutskever, I., & Hinton, G. (2007). ImageNet Classification with Deep Convolutional Neural Networks. *Handbook of Approximation Algorithms and Metaheuristics*, 1–1432. <https://doi.org/10.1201/9781420010749>
7. Loncomilla, P. (2016). Deep learning: Redes convolucionales. Recuperado de <https://ccc.inaoep.mx/~pgomez/deep/presentations>.
8. Moreno, A. (2019). Autor: Álvaro Artola Moreno Tutor: José Antonio Pérez Carrasco.

9. Monroy, J, Ramírez, A, Alejo, R y López, E.(2017) Aspectos relevantes para mejorar el desempeño del algoritmo Backpropagation. *Revista de Tecnologías de la Información* 2017, 4-13: 1-11
10. Nayak, S. (2018). Understanding AlexNet | Learn OpenCV. Retrieved December 2, 2019, from <https://www.learnopencv.com/understanding-alexnet/>
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115(3), 211–252. <https://doi.org/10.1007/s11263-015-0816-y>
12. OMS.(2018).Air-pollution-INFOGRAPHICS-English-4-1200px.jpg (1200×1200). [https://doi.org/CLEAN AIR FOR HEALTH](https://doi.org/CLEAN_AIR_FOR_HEALTH)
13. Zhu, J., Zhang, E. & Del Rio-Tsonis, K., 2001. Eye Anatomy. e LS.

References

1. Deepthi Jain, B., Thakur, S. M., & Suresh, K. V. (2018). Visual Assistance for Blind Using Image Processing. *Proceedings of the 2018 IEEE International Conference on Communication and Signal Processing, ICCSP 2018*, 499–503. <https://doi.org/10.1109/ICCSP.2018.8524251>
2. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... Dean, J. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25 (1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning (Adaptive Computation and Machine learning)*. Adaptive Computation and Machine Learning Series. <https://doi.org/10.1016/B978-0-12-391420-0.09987-X>
4. Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation* (1st ed.). Upper Saddle River, NJ, USA: Prentice Hall PTR.
5. Kacorri, H., Kitani, K. M., Bigham, J. P., & Asakawa, C. (2017). People with visual impairment training personal object recognizers: Feasibility and challenges. *Conference on Human Factors in Computing Systems - Proceedings, 2017-May*, 5839–5849. <https://doi.org/10.1145/3025453.3025899>

6. Krizhevsky, A., Sutskever, I., & Hinton, G. (2007). ImageNet Classification with Deep Convolutional Neural Networks. *Handbook of Approximation Algorithms and Metaheuristics*, 1–1432. <https://doi.org/10.1201/9781420010749>
7. Loncomilla, P. (2016). Deep learning: Convolutional networks. Recovered from <https://ccc.inaoep.mx/~pgomez/deep/presentations>.
8. Moreno, A. (2019). Author: Álvaro Artola Moreno Tutor: José Antonio Pérez Carrasco.
9. Monroy, J, Ramírez, A, Alejo, R and López, E. (2017) Relevant aspects to improve the performance of the Backpropagation algorithm. *Information Technology Magazine* 2017, 4-13: 1-11
10. Nayak, S. (2018). Understanding AlexNet | Learn OpenCV. Retrieved December 2, 2019, from <https://www.learnopencv.com/understanding-alexnet/>
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S.,... Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. *International Journal of Computer Vision*, 115 (3), 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
12. WHO. (2018) .Air-pollution-INFOGRAPHICS-English-4-1200px.jpg (1200 × 1200). [https://doi.org/CLEAN AIR FOR HEALTH](https://doi.org/CLEAN_AIR_FOR_HEALTH)
13. Zhu, J., Zhang, E. & Del Rio-Tsonis, K., 2001. Eye Anatomy. and LS.

Referências

1. Deepthi Jain, B., Thakur, S.M. & Suresh, K.V. (2018). Assistência visual para cegos usando o processamento de imagens. *Anais da Conferência Internacional do IEEE de 2018 sobre Comunicação e Processamento de Sinais, ICCSP 2018*, 499–503. <https://doi.org/10.1109/ICCSP.2018.8524251>
2. Esteva, A., Robicquet, A., Ramsundar, B., Kuleshov, V., DePristo, M., Chou, K., ... Dean, J. (2019). Um guia para a aprendizagem profunda em saúde. *Nature Medicine*, 25 (1), 24-29. <https://doi.org/10.1038/s41591-018-0316-z>
3. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Aprendizagem Profunda (Computação Adaptativa e Aprendizado de Máquina)*. Série de Computação Adaptativa e Aprendizado de Máquina. <https://doi.org/10.1016/B978-0-12-391420-0.09987-X>
4. Haykin, S. (1994). *Redes Neurais: Uma Fundação Abrangente (1ª ed.)*. Upper Saddle River, NJ, EUA: Prentice Hall PTR.

5. Kacorri, H., Kitani, K.M., Bigham, J.P. e Asakawa, C. (2017). Pessoas com deficiência visual treinando reconhedores de objetos pessoais: Viabilidade e desafios. Conferência sobre fatores humanos em sistemas de computação - Proceedings, 2017-maio, 5839-5849. <https://doi.org/10.1145/3025453.3025899>
6. Krizhevsky, A., Sutskever, I., & Hinton, G. (2007). Classificação ImageNet com redes convolucionais profundas. Manual de Algoritmos de Aproximação e Metaheurísticas, 1–1432. <https://doi.org/10.1201/9781420010749>
7. Loncomilla, P. (2016). Aprendizagem profunda: redes convolucionais. Recuperado de <https://ccc.inaoep.mx/~pgomez/deep/apresentações>.
8. Moreno, A. (2019). Autor: Álvaro Artola Moreno Tutor: José Antonio Pérez Carrasco.
9. Monroy, J, Ramírez, A, Alejo, R e López, E. (2017) Aspectos relevantes para melhorar o desempenho do algoritmo de retropropagação. Revista de Tecnologia da Informação 2017, 4-13: 1-11
10. Nayak, S. (2018). Entendendo AlexNet | Aprenda OpenCV. Recuperado em 2 de dezembro de 2019, em <https://www.learnopencv.com/understanding-alexnet/>
11. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... Fei-Fei, L. (2015). Desafio de reconhecimento visual em grande escala do ImageNet. International Journal of Computer Vision, 115 (3), 211-252. <https://doi.org/10.1007/s11263-015-0816-y>
12. OMS (2018) .A poluição do ar-INFOGRAPHICS-Português-4-1200px.jpg (1200 × 1200). [https://doi.org/CLEAN AIR FOR HEALTH](https://doi.org/CLEAN_AIR_FOR_HEALTH)
13. Zhu, J., Zhang, E. & Del Rio-Tsonis, K., 2001. Eye Anatomy. e LS.

©2019 por los autores. Este artículo es de acceso abierto y distribuido según los términos y condiciones de la licencia Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0) (<https://creativecommons.org/licenses/by-nc-sa/4.0/>).