

# Confección de dataset de patentes de automóviles argentinos para entrenamiento de redes neuronales.

Ferreya Biron, Martín, Maidana, Carlos Eduardo

Departamento de Ingeniería e Investigaciones Tecnológicas, Universidad Nacional de La Matanza, Buenos Aires, Argentina.

{mferreyra,cmaidana} @unlam.edu.ar

## RESUMEN

En el marco de una investigación más extensa, se conformó un vasto dataset compuesto por imágenes de escenas donde se encuentran automóviles, imágenes de automóviles, patentes, caracteres y dígitos de las mismas en ambas versiones de la patente argentina. Dicho dataset fue utilizado para entrenar distintas redes neuronales y depurado de diversas formas en varias ocasiones. El mismo posee más de 297.000 imágenes y puede ser utilizado en una investigación de posgrado o doctorado (con la previa anuencia de la UNLaM) para entrenar una red neuronal o para cualquier otro tipo de investigación en donde se necesiten este tipo de datos. En este trabajo se muestran las características de dicho dataset y los resultados al utilizarlo en ciertas redes neuronales.

**Palabras Clave:** Dataset, Redes neuronales, Inteligencia artificial, Chapa patente argentina.

## CONTEXTO

El presente documento es resultado de la investigación “*Obtención de dataset a partir de imágenes capturadas para su utilización en sistema experto de reconocimiento de imágenes*” la cual es una continuación de la investigación anterior “*Obtención de dataset a partir de imágenes*

*capturadas del estacionamiento de UNLaM para su utilización en sistema experto de reconocimiento de imágenes*” inconclusa por distintas desavenencias y la imposibilidad de llevarla a cabo debido a las diversas situaciones ocasionadas por la pandemia de COVID-19.

## 1. INTRODUCCIÓN

En la situación en la que se encuentran desarrolladas las redes neuronales hoy en día, es de absoluta y vital importancia tener un nutrido dataset, bien clasificado y depurado para poder entrenar correctamente una red neuronal [1]. Y el hincapié de hoy en día está dado porque una red neuronal va a ser más precisa si la cantidad de datos utilizados en esta etapa es verdaderamente abundante. Es cierto que existen técnicas como *data augmentation* [2] que permiten mitigar una deficiencia de esta materia prima, pero sin embargo poseer un importante dataset, cuantioso y variado hace a un mejor desempeño a la hora de utilizar una red neuronal. También es una verdad insoslayable que existen diversos datasets que cubren, en general, una gran variedad de objetos que pueden ir desde números, como la del famoso dataset MNIST [3], pasando por datasets de rostros humanos como DigiFace-1M [4] hasta llegar a datasets de mariposas como Leeds Butterfly Dataset [5] e incluso estos

datasets pueden contener varias clases de objetos como Imagenet [6] o COCO [7]. La dificultad comienza cuando los datasets que se necesitan no son de un objeto general, como un carácter escrito o como un rostro humano, sino algo más local o específico de una región, como por ejemplo un dataset de imágenes de patentes argentinas. En estos casos, desde nuestra perspectiva, no es una opción utilizar datos sintéticos [9].

## 2. LÍNEA DE INVESTIGACIÓN Y DESARROLLO

Al anhelar poseer una forma de confirmar qué automóviles ingresaban en el estacionamiento universitario y también, por otras razones se propuso entrenar un sistema experto con el fin de poder detectar cuales eran las patentes de los automóviles que ingresaba al mismo. Pero nos encontramos con el inconveniente de no tener a nuestra disposición una base de datos de imágenes con patentes argentinas. Debido a esto llevamos a cabo la primera investigación mencionada en la sección Contexto, la cual no pudo finalizarse por los motivos ya explicados. En esta investigación, en líneas generales, se planteaba obtener un dataset de imágenes de patentes argentinas, para su posterior utilización en el entrenamiento de un sistema experto. Al no poder cumplirse completamente con los objetivos de la misma, se continuó la investigación con la siguiente: *“Obtención de dataset a partir de imágenes capturadas para su utilización en sistema experto de reconocimiento de imágenes”*. En esta se propone completar el dataset que se deseaba confeccionar en la investigación anterior y utilizarlo en el entrenamiento de diversas redes neuronales. Este documento presenta las características del dataset obtenido, y a modo de ejemplo cómo se desempeñó en el

entrenamiento de una red neuronal YoloV3-Tiny[8].

Para elaborar este dataset se propuso obtener al menos 100 horas de grabación de tráfico, intentado que estas fueran lo más heterogéneas posibles. Luego de haber obtenido estas, se revisaron en su totalidad en búsqueda de vehículos, patentes del tipo 1994 y 2015 y sus respectivos caracteres. Estas búsquedas persiguieron el fin de obtener y resguardar la información de las posiciones de los elementos encontrados en la imagen, su alto, su ancho y tipo de objeto, entre otros sumados a la imagen del objeto. Además se conservaron los fotogramas donde se encuentran cada una de las imágenes recolectadas. Para este fin fue desarrollado un software interno que ayudó en gran proporción a realizar esta ardua tarea, haciéndola más tolerable, además de automatizar el resguardo de estos datos. El acopio de la información se realizó en archivos de texto plano que contienen estos valores separados por comas (intentando hacer más fácil a futuro su manejo). Los datos de los objetos obtenidos se encuentran divididos en tres archivos que poseen estructuras distintas dependiendo el caso.

Luego de una extensa comprobación de los datos, valiéndonos de algunas automatizaciones programadas por nosotros y de varias inspecciones oculares pudimos concluir la creación del dataset y realizar los cálculos estadísticos.

Con el dataset concluido, nos dispusimos a desarrollar un sistema que fuese capaz, a partir del uso de redes neuronales y del dataset construido, de detectar vehículos en una escena, luego su patente y finalmente los caracteres que la conforman. De esta forma, si realmente el dataset confeccionado es idóneo, las redes

neuronales entrenadas deberán ser capaces de realizar las detecciones pertinentes.

Para cumplir dichos objetivos consideramos la red neuronal Yolo V3 , en su versión tiny, para la detección de vehículos y patentes. Respecto a los caracteres, luego de numerosos intentos desistimos de utilizar Yolo para el reconocimiento de los mismos decantándonos por la red neuronal Xception. En cambio utilizamos Yolo para la localización de los caracteres en la patente.

En resumen el pipeline propuesto es el siguiente:

1. Captura de la escena
2. Detección del vehículo (Yolo)
3. Localización y categorización de la patente. (Yolo)
4. Localización de los caracteres en la patente (Yolo) (Se utilizan dos redes neuronales para esto , una para las patentes 1994 y otra para las patentes 2015)
5. Reconocimiento de los caracteres y dígitos (Xception)

### **3. RESULTADOS OBTENIDOS/ESPERADOS**

Los primeros resultados a destacar son las estadísticas del dataset confeccionado. Como aseveramos en el apartado anterior, la cantidad de horas de vídeo conseguidas asciende a la cifra de 100 horas en distintos lugares como las localidades de Morón, San Justo, Villa Luzuriaga, Ramos Mejía, Haedo, Ituzaingó y CABA en las cuales pudimos obtener **29.278 imágenes de vehículos, 33.829 imágenes de patentes** , donde **11.301 corresponden a patentes 2015** y **22.528 corresponden a patentes 1994**. Cabe

mencionar que no todas las patentes tienen una imagen de vehículo asociado y no toda imagen de vehículo obtenido posee una imagen de una patente asociada.

Finalmente de las patentes recolectadas pudimos obtener **62.810 caracteres y 62.958 dígitos de patentes 1994** y **42.704 caracteres y 32.028 dígitos de patentes 2015**. Todas las imágenes se encuentran en las más variadas condiciones meteorológicas, tamaños y perspectivas. La distribución de los datos obtenidos se puede consultar en [10]

Respecto al entrenamiento de las redes neuronales, tuvimos en general una respuesta positiva al entrenar cada una de ellas, más allá de los inconvenientes con los que nos pudimos encontrar a lo largo de esta tarea. En todos los casos utilizamos redes neuronales pre entrenadas intentando hacer uso del concepto de *transferencia de aprendizaje* [2]. Las razones de esto son diversas , aunque la principal es la gran cantidad de tiempo que se requiere para entrenar correctamente estas grandes redes y la segunda es la posibilidad de caer rápidamente en *overfitting* debido a las características de estas.

La primera red neuronal que entrenamos fue la correspondiente a la localización de una patente en la imagen de un vehículo y su categorización en una patente del tipo 1994 o 2015. Para realizar dicha tarea se entrenó una red neuronal YoloV3-Tiny con 24.223 imágenes de vehículos en los cuales se podían observar las patentes. De este conjunto de datos se utilizó un 70% para entrenamiento y un 30% para testeo. Estas imágenes fueron convertidas a escala de grises. Luego de aproximadamente medio millón de epochs, comprobamos el desempeño en el conjunto de entrenamiento y en el conjunto de testeo,

obteniéndose un **porcentaje de acierto del 95%**.

Después de haber concluido con el entrenamiento anterior nos enfocamos en otras dos redes neuronales YoloV3-Tiny con el fin de hallar la ubicación de los caracteres y dígitos. Para el caso de los caracteres de las patentes 2015 tomamos aproximadamente todas las patentes 2015 mencionadas en este trabajo, convertimos estas a escala de grises y dividimos el conjunto de datos en tres, un conjunto de validación con el 10% de las imágenes totales, un conjunto de imágenes de entrenamiento conformado por el 70% del 90% del total de imágenes, y un conjunto de testeo conformado por el 30% del 90% del total de imágenes. Entrenamos esta red neuronal con 1 millón de epochs, modificando aleatoriamente los conjuntos de entrenamiento y testeo cada 50000 epochs. Al finalizar comprobamos el desempeño de la red neuronal con el conjunto de validación (que nunca se utilizó en el entrenamiento) obteniéndose un **porcentaje de acierto del 79%**.

Del mismo modo y aproximadamente con la misma cantidad de epochs entrenamos otra red neuronal Yolo para poder detectar la localización de los caracteres y dígitos en las patentes 1994. **Obtuvimos un desempeño en el set de validación del 73,73%**

Continuando con la validación del dataset nos dedicamos a entrenar dos redes neuronales Xception, una de ellas dedicada a reconocer los caracteres de la patente y la otra dedicada a reconocer dígitos de las mismas en ambos casos sin importar la variante de la patente.

Para el caso de la red neuronal destinada a reconocer caracteres, eliminamos las capas finales de esta red neuronal, y colocamos 3 capas de MLP

totalmente conectados de 800 perceptrones junto con una capa de salida de 26 perceptrones. Estas capas, junto con las dos últimas convolucionales fueron las que entrenamos durante 40 epochs, de una forma similar a como lo hicimos con la red neuronal Yolo para la localización de caracteres y dígitos. Al finalizar el entrenamiento, pusimos a prueba esta red con el conjunto de datos de validación (que no conformó parte del entrenamiento) y obtuvimos un **95,11% de acierto en el reconocimiento de los caracteres**.

Para el caso de la red neuronal destinada a reconocer dígitos, eliminamos las capas finales y colocamos 3 MLP de 800 perceptrones junto con una capa de salida de 10 perceptrones. Estas capas, unidas con la última convolucional fueron las entrenadas durante 120 epochs. El entrenamiento fue realizado de la misma forma que la red neuronal anterior. Al finalizar el entrenamiento y comprobar el rendimiento de esta red conseguimos un **97,79% de acierto en el reconocimiento de dígitos** sobre el set de datos correspondientes a la validación.

Por último y no menos importante entrenamos una red YoloV3-Tiny para detectar los vehículos que aparecen en una escena. Para esto aplicamos el mismo método de entrenamiento explicado anteriormente para la localización de las patentes. Utilizamos el 70% de las imágenes de los vehículos para entrenamiento y el 30% restante como imágenes de testeo. Luego de aproximadamente medio millón de epochs obtuvimos un **92% de efectividad** al poner a prueba esta red con el total de las imágenes de vehículos del dataset.

Si bien los datos mencionados anteriormente son alentadores, en las pruebas realizadas con imágenes que no

corresponden a ninguno de los sets de datos utilizados , no obtuvimos un buen desempeño en algunos casos, ocurriendo esto principalmente en el reconocimiento de caracteres y de dígitos y de vehículos.

Respecto a la detección de vehículos, cometimos un error al capturarlos. No todos los vehículos en las escenas fueron recolectados, solamente aquellos en los cuales se podía observar la patente. Esto produjo un inconveniente al entrenar la red neuronal, ya que en el entrenamiento se ofrecían datos confusos: en una escena existían vehículos , que no fueron marcados como vehículos, y algunos vehículos que sí lo fueron.

Por otra parte, en el caso de los caracteres y dígitos , el reconocimiento no era aceptable. Pocas patentes eran reconocidas correctamente. Entendemos que esto se debía a una incompleta segmentación de los caracteres y dígitos, solucionando dicho inconveniente redimensionando la imagen de las patentes cinco veces, antes de la etapa de segmentación. De esta forma la misma era más precisa y los caracteres podían ser reconocidos con más certeza.

#### 4 .FORMACIÓN DE RECURSOS HUMANOS

Los integrantes de este desarrollo forman parte del plantel docente e investigadores del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de la Matanza. En conjunto, los integrantes de este grupo

poseen conocimientos de electrónica e informática.

#### 6. Bibliografía

- [1]S. Raschka y V. Mirjalili, *Python machine learning: machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*, Third edition. Birmingham Mumbai: Packt, 2019.
- [2] U. Michelucci, *Advanced applied deep learning: convolutional neural networks and object detection*. New York: Apress, 2019.
- [3]Yann Lecun, “The MNIST database of handwritten digits”, <https://www.clarin.com>
- [4]G. Bae *et al.*, “DigiFace-1M: 1 Million Digital Face Images for Face Recognition”. arXiv, doi: [10.48550/arXiv.2210.02579](https://arxiv.org/abs/10.48550/arXiv.2210.02579).
- [5]Wang, Josiah, “Leeds Butterfly Dataset”. Zenodo, doi: [10.5281/ZENODO.7559420](https://zenodo.org/record/10.5281/ZENODO.7559420).
- [6]“ImageNet”. <https://www.image-net.org/>
- [7]“COCO - Common Objects in Context”. <https://cocodataset.org/#home>
- [8] J. Redmon y A. Farhadi, “YOLOv3: An Incremental Improvement”. arXiv, 8 de abril de 2018. doi: [10.48550/arXiv.1804.02767](https://arxiv.org/abs/10.48550/arXiv.1804.02767).
- [9] “Synthetic data”, *Wikipedia.*: [https://en.wikipedia.org/w/index.php?title=Synthetic\\_data&oldid=112992158](https://en.wikipedia.org/w/index.php?title=Synthetic_data&oldid=112992158)
- [10]<https://github.com/mfcategrep/DatasetWicc2023>